Studying Strategic Thinking
by Monitoring Search for Hidden Payoff Information

Miguel A. Costa-Gomes, University of Aberdeen

And

Vincent P. Crawford, University of California, San Diego

Cemmap/ELSE Workshop

“Experimental Analysis of Procedural Rationality in Games and Decisions”

University College London, 4 March 2009

Introduction

This talk concerns experiments that study strategic thinking by eliciting subjects' initial responses to series of different but related games, while monitoring and analyzing the patterns of subjects' searches for hidden but freely accessible payoff information along with their decisions.

The talk is based on three papers:

Costa-Gomes and Crawford, "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review* 2006 ("CGC").

Crawford, "Look-ups as the Windows of the Strategic Soul: Studying Cognition via Information Search in Game Experiments," in Andrew Caplin and Andrew Schotter, editors, *Perspectives on the Future of Economics: Positive and Normative Foundations*, Volume 1, Handbooks of Economic Methodologies, Oxford University Press, 2008

Costa-Gomes and Crawford, "Studying Cognition via Information Search in Two-Person Guessing Game Experiments," still in preparation.

Other experiments that study strategic thinking via search patterns

Camerer, Johnson, Rymon, and Sen, "Cognition and Framing in Sequential Bargaining for Gains and Losses," in Kenneth Binmore, Alan Kirman, and Piero Tani, editors, *Frontiers of Game Theory*, 1993 ("CJ")

Johnson, Camerer, Sen, and Rymon, "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining," *Journal of Economic Theory* 2002 ("CJ")

Costa-Gomes, Crawford, and Broseta, "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica* 2001 ("CGCB")

Camerer and Johnson, "Thinking about Attention in Games: Backward and Forward Induction," in Isabel Brocas and Juan Carrillo (editors), *The Psychology of Economic Decisions, Volume Two: Reasons and Choices*, Oxford, 2004

Wang, Spezio, and Camerer, "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation To Understand Truth-telling and Deception in Games," 2008

Experiments that study strategic thinking via search durations

Ariel Rubinstein, "Instinctive and Cognitive Reasoning: A Study of Response Times," *Economic Journal* 2007

Adapting methods introduced to the experimental game theory literature by CJ and CGCB—previously used extensively to analyze decisions, for example by Payne, Bettman, and Johnson 1993—CGC elicited subjects' initial responses to a series of 16 two-person guessing games designed for this purpose, while monitoring and analyzing the patterns of subjects' searches for hidden but freely accessible payoff information.

Following CGCB, CGC then used an explicit, procedurally rational model of cognition to analyze subjects' searches along with their decisions.

GCG’s analysis shows that with careful design, subjects’ search patterns can sometimes directly reveal the algorithms used to choose their decisions, in such cases making it possible to identify subjects’ decision rules even without observing their decisions.

The analysis also shows that decisions and search are complementary, together making it possible to identify subjects’ decision rules more precisely than would be possible even with unlimited decision data.

CGC’s analysis also illustrates some novel analytical and econometric issues that arise in analyzing process data.

Motivation

The topic of studying strategic thinking via information search raises two questions of motivation:

- Why study strategic thinking when even unthinking people are likely eventually to converge to equilibrium anyway?

- Why study strategic thinking by monitoring and analyzing process data if the goal is only to predict decisions?

Why study strategic thinking?

Strategic thinking is an essential part of human interaction, but one whose importance from a behavioral point of view has been downplayed.

Most applications of game theory in economics and game theory rely on Nash equilibrium.

But while equilibrium can be viewed as a model of strategic thinking, there are many applications for which it is not an adequate model of behavior.

Players' strategies will be in equilibrium if they are rational and have the same beliefs about each other's strategies.

Accepting rationality for the sake of argument, there are two possible justifications for the assumption that players have the same beliefs:

- Thinking: If players have perfect models of each other's decisions, strategic thinking will lead them to have the same beliefs immediately, and so play an equilibrium even in their initial responses to a game.

- Learning: Even without perfect models, if players repeatedly play analogous games, experience may eventually allow them to predict each others' decisions well enough to play an equilibrium in the limit.

In many applications the theoretical conditions for learning to converge to equilibrium are approximately satisfied, and in such settings both experimental and field evidence tends to support assuming that steady-state strategy choices are in equilibrium (with some qualifications).

In applications where only long-run outcomes matter, or where equilibrium is unique, or where equilibrium selection does not depend on the details of learning, analysis can safely rely entirely on equilibrium.

However, many other applications involve games played without clear precedents, so that the standard learning justification for equilibrium is unavailable.

In other applications, eventual convergence to equilibrium is assured, but initial as well as limiting outcomes matter (e.g. FCC Spectrum auction).

And in still other applications, convergence is assured and only long-run outcomes matter, but the equilibrium is selected from multiple possibilities via history-dependent learning dynamics.

All such applications depend on reliably predicting initial responses to games, which may require a non-equilibrium model of strategic thinking.

As will be seen, empirically successful models of strategic thinking normally allow equilibrium behavior, but do not assume equilibrium in all games.

Instead they assume that players follow strategic but non-equilibrium decision rules, which yield decisions that mimic equilibrium in simple games, but may deviate systematically in more complex games.

The models thereby provide a way to predict, in a given game, whether players’ responses are likely to deviate from equilibrium, and if so, how.

Why study process data?

An experimental design could, in principle, separate the decisions implied by different kinds of strategic thinking well enough to allow us to infer thinking entirely from decisions.

But in economically interesting games, our ability to distinguish among models of strategic thinking is near the limits of experimental feasibility.

For example, although CGC's design, described below, is quite powerful from the standpoint of studying decisions alone, it leaves open some important questions regarding subjects' decision rules.

If decision data were free, it might be optimal to address open questions just by gathering more decision data, perhaps in new environments.

But decision data are far from free, and existing methods for gathering them are fairly easily adapted to gather process data at the same time.

Further, with careful design, monitoring search for hidden payoff information can give us an independent "take" on strategic thinking, one that is more directly related to cognition than are decisions.

As will be seen, monitoring search sometimes allows us to directly observe the algorithms subjects use to make their decisions, and to distinguish mistakes from intended behavior.

Thus, exclusive reliance on gathering more decision data seems unlikely always to be optimal: At least for studying thinking, good research strategies should be open to process as well as decision data, even if this requires developing new methods of analysis.

Outline of the talk

The talk begins by summarizing CGC's experimental design.

It then discusses CGC's results for subjects' decisions, introducing the model based on strategic thinking "types" that underlies their analysis and highlighting econometric issues that remain open.

It next raises some questions regarding subjects' thinking that are not adequately resolved by analyzing decisions alone, but which might be resolved by analyzing decisions and information search.

The talk then turns to CGC’s analysis of cognition and search.

The types used to analyze decisions play an essential role in analyzing search.

CGC’s model of cognition and search takes a procedural view of decision-making:

In a given game, a subject’s type first determines his search, and his type and search then jointly determine his decision.

In the analysis, the types provide a basis for the enormous space of possible decision and search sequences, imposing enough structure to allow us to describe subjects’ behavior in a comprehensible way and to make it meaningful to ask how their decisions and searches are related.

The talk concludes by summarizing CGC’s results for information search and highlighting open econometric issues involving search.

CGC’s experimental design

CGC’s experiments randomly and anonymously paired subjects to play a series of two-person guessing games, with no feedback between games.

The design suppresses learning from experience and repeated-game effects in order to elicit subjects’ initial responses, game by game.

The goal is to focus on how people model others’ decisions by studying strategic thinking “uncontaminated” by learning from experience.

“Eureka!” learning remains possible, but CGC tested for it and found it to be rare.

(The results yield insights into cognition that also help us think about how to model learning from experience, but that’s another story.)

CGC's design combines the variation of the games each subject played of CJ's 1993 design and Stahl and Wilson's 1995 *GEB* design with the very large strategy spaces of Nagel's 1995 *AER* and Ho, Camerer, and Weigelt's ("HCW") 1998 *AER* designs.

This combination greatly enhances the design's power:

A subject's profile of guesses forms a "fingerprint" that identifies his strategic thinking more precisely than is possible by observing his responses to a series of different games with small strategy spaces or any single game, even with a very large strategy space.

In CGC’s two-person guessing games, each player has a lower and an upper limit, both strictly positive, each taking one of two possible values.

However, players are not required to guess between their limits: Instead guesses outside the limits are automatically adjusted up to the lower or down to the upper limit as necessary—a trick to enhance the separation of decision rules via their information search implications.

Each player also has his own target, taking one of four possible values.

A player’s payoff increases with the closeness of his adjusted guess to his target times the other player’s adjusted guess.

The targets and limits vary independently across players and 16 games, with targets either both less than one, both greater than one, or “mixed”.

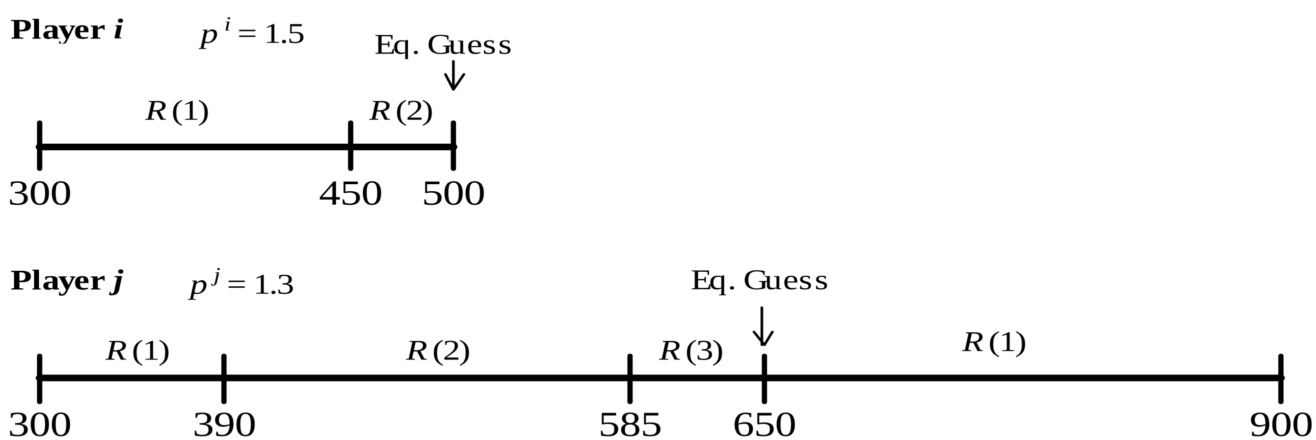
(In Nagel’s and HCW’s previous guessing game experiments, the targets and limits were always the same for both players, and they varied only across treatments with different subject groups, or not at all.)

For example, in game γ4δ3 (#5 in CGC's Table 3), player *i*'s limits and target are [300, 500] and 1.5; and player *j*'s are [300, 900] and 1.3.

The product of targets 1.5 × 1.3 > 1, and players' equilibrium adjusted guesses are determined (not always directly) by their upper limits:

i's equilibrium adjusted guess equals his upper limit of 500, but *j*'s is below his upper limit at 650.

In the figure, guesses in the interval *R*(*k*) are eliminated in round *k* of iterated dominance; thus the game is finitely dominance solvable.



CGC’s sixteen games are all finitely dominance-solvable, in from 3 to 52 rounds, with essentially (due to automatic adjustment) unique equilibria.

The way in which equilibrium is determined in game γ4δ3, by players’ upper limits (in the indirect sense illustrated in the example) when the product of their targets is greater than 1—or by their lower limits when the product is less than 1—is general in CGC’s games.

CGC’s design exploits the discontinuity of the equilibrium correspondence when the product of targets is 1 by including some games that differ mainly in whether the product is slightly greater, or slightly less, than 1.

Equilibrium responds strongly to such differences, but empirically plausible non-equilibrium decision rules are largely unmoved by them.

That equilibrium is jointly determined by both players’ payoff parameters also helps to separate search implications of equilibrium and other rules.

CGC’s types-based model of decisions

Following CGCB and other previous work in this area, CGC’s analysis of decisions uses a types-based structural non-equilibrium model.

The model assumes that each subject’s guesses are determined in all 16 games, up to logit errors, by a single decision rule or “type” (as they are called in this literature; no relation to private-information variables).

CGC’s types, listed on the next slide, all build in risk-neutrality and rule out social preferences, again following previous work.

Risk aversion and social preferences are somewhat implausible in this context, and the results and CGC’s specification test, explained below, suggest that they were not important factors in subjects’ decisions.

The list of types also excludes some others that might seem plausible, mainly because they did not show up significantly in earlier analyses; CGC’s specification test doesn’t find any empirically important omissions.

- *L0*, *L1*, *L2*, and *L3*, with *L0* uniform random between a player's limits, *L1* best responding to *L0*, *L2* to *L1*, and so on.

  (*L0* represents a subject's instinctive, nonstrategic reaction to the game, and usually has zero estimated population frequency. *Lk* for *k* > 0 is rational, but deviates from equilibrium because it uses a simplified model of others' decisions. It is *k*-rationalizable, and so coincides with equilibrium in games that are *k*-dominance solvable.)

- *D1* and *D2*, which does one round (respectively, two) of iterated dominance and best responds to a uniform prior over its partner's remaining decisions (a selection from the *k*-rationalizable strategies).

  (By a quirk of our notation, *L2* is *D1*'s cousin, and *L3* is *D2*'s. Those pairs' guesses are perfectly confounded in Nagel's *AER* 1995 games; and in two-person games *Lk* guesses are *k*-rationalizable, like *Dk-1*'s.)

- *Equilibrium*, which makes its equilibrium decisions.

- *Sophisticated*, which best responds to the probabilities of others' decisions, proxied by subjects' observed frequencies. (*Sophisticated* is an ideal, included to learn if any subjects have an understanding of others' decisions that transcends mechanical rules.)

CGC's results for decisions

The large strategy spaces of CGC's games and their variation of targets and limits greatly enhance the separation of types' implications.

(In the table, a player's lower limit, upper limit, and target are denoted a_i,b_i, and p_i respectively; and his partner's are denoted a_j,b_j, and p_j.)

Types' guesses in the 16 games, in (randomized) order played													
Game	a_i	b_i	p_i	a_j	b_j	p_j	***L1***	***L2***	***L3***	***D1***	***D2***	***Eq***	***So***
1	100	900	1.5	300	500	0.7	600	525	630	600	611.25	750	630
2	300	900	1.3	300	500	1.5	520	650	650	617.5	650	650	650
3	300	900	1.3	300	900	1.3	780	900	900	838.5	900	900	900
4	300	900	0.7	100	900	1.3	350	546	318.5	451.5	423.15	300	420
5	100	500	1.5	100	500	0.7	450	315	472.5	337.5	341.25	500	375
6	100	500	0.7	100	900	0.5	350	105	122.5	122.5	122.5	100	122
7	100	500	0.7	100	500	1.5	210	315	220.5	227.5	227.5	350	262
8	300	500	0.7	100	900	1.5	350	420	367.5	420	420	500	420
9	300	500	1.5	300	900	1.3	500	500	500	500	500	500	500
10	300	500	0.7	100	900	0.5	350	300	300	300	300	300	300
11	100	500	1.5	100	900	0.5	500	225	375	262.5	262.5	150	300
12	300	900	1.3	300	900	1.3	780	900	900	838.5	900	900	900
13	100	900	1.3	300	900	0.7	780	455	709.8	604.5	604.5	390	695
14	100	900	0.5	300	500	0.7	200	175	150	200	150	150	162
15	100	900	0.5	100	500	0.7	150	175	100	150	100	100	132
16	100	900	0.5	100	500	1.5	150	250	112.5	162.5	131.25	100	187

Of the 88 subjects in CGC's main treatments, 43 made guesses that complied *exactly* (within 0.5) with one type's guesses in from 7 to 16 of the games (20 *L1*, 12 *L2*, 3 *L3*, and 8 *Equilibrium*).

For example, CGC's Figure 2 (next slide) shows the strategic thinking "fingerprints" of the twelve subjects whose guesses conformed very closely (that is, with high rates of exact compliance) to *L2*'s guesses.

72% (138) of these subjects' 192 guesses were exact *L2* guesses; only their deviations are shown in Figure 2.

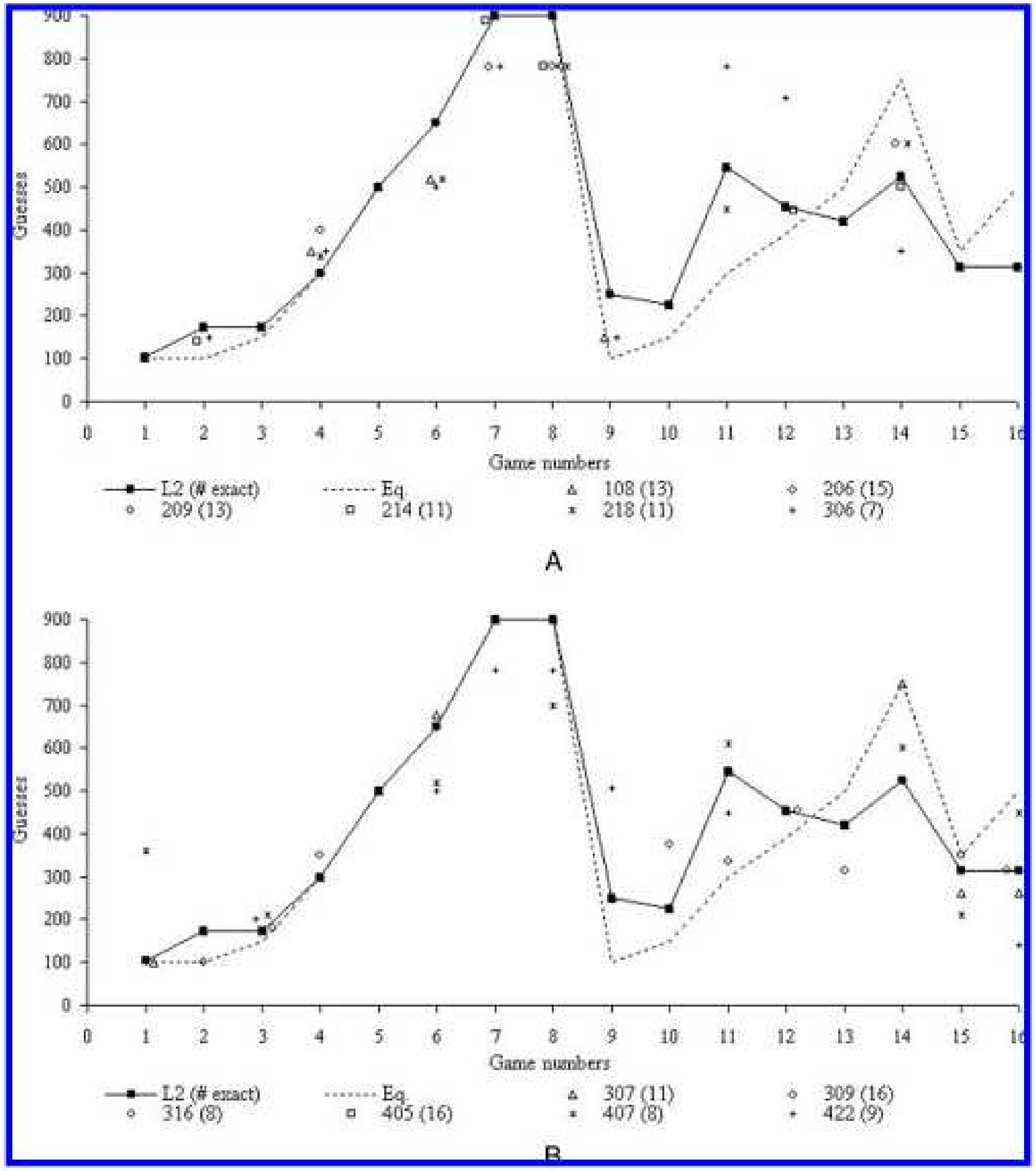


FIGURE 2. "FINGERPRINTS" OF 12 APPARENT *L2* SUBJECTS

Notes: Only deviations from *L2*'s guesses are shown. Of these subjects' 192 guesses, 138 (72 percent) were exact *L2* guesses.

Given how strongly CGC’s design separates types’ guesses, and that guesses could take from 200 to 800 different rounded values in the games, these subjects’ exact compliance rates are far higher than could possibly occur by chance:

If a subject chooses 525, 650, 900 in games 1-3, both intuitively and econometrically we already “know” he’s an *L2*.

Further, because CGC’s definition of *L2* builds in risk-neutral, self-interested rationality, we also know that with such high exact compliance, a non-*Equilibrium* subject’s deviations from equilibrium are “caused” not by irrationality, risk aversion, altruism, spite, or confusion, but by his simplified (in this case *L1*) model of others.

Guessmetrics

CGC's other 45 subjects made guesses that conformed exactly to one of the types less frequently; analyzing their guesses requires econometrics.

Our econometric approach builds on Harless and Camerer 1994 *Econometrica*, El-Gamal and Grether 1995 *JASA*, Stahl and Wilson 1994 *JEBO* and 1995 *GEB*, and CGCB; but we estimate subject by subject, and because of the very high sample frequency of exact guesses, we use a maximum-likelihood error-rate model with "spike-logit" errors:

We assume that in each game, a subject makes his type's guess exactly (within 0.5) with probability 1- ε and otherwise makes logit errors; this gives extra likelihood credit for exact guesses, whose likelihood weight is discontinuously higher than guesses that are close but not within 0.5.

Estimating a mixture model as in CGCB and most other previous studies is often theoretically superior; but in an exploratory study of cognition, estimating subject by subject is safer and, comparing CGCB with subject-by-subject estimates in its earliest version, likely yields similar estimates.

Subject *i*'s log-likelihood for guesses reduces to:

(7) $$(G-n^{ik})\ln(G-n^{ik})+n^{ik}\ln(n^{ik})+\sum_{g\in N^{ik}}\ln d_g^k(R_g^i(x_g^i),\lambda)-G\ln G,$$

where *g* indexes games and *k* types.

The first two terms concern exact guesses; $d_g^k(R_g^i(x_g^i),\lambda)$ is the standard logit term for non-exact guesses, with deviation costs measured using each type's beliefs; and *λ* is the logit precision.

The maximum likelihood estimate of ε is n^{ik} /*G*, the sample frequency of subject *i*'s non-exact guesses for type *k*.

The maximum likelihood estimate of *λ* is the standard logit precision, restricted to non-exact guesses.

The maximum likelihood estimate of the subject's type *k* maximizes (7) over *k*, given the estimated ε and *λ*, trading off the count of exact guesses against the logit cost of deviations.

Estimation yields type estimates as in column 3 of Table 1: 43 *L1*, 20 *L2*, 3 *L3*, 5 *D1*, 14 *Equilibrium*, and 3 *Sophisticated*.

(Some of these estimates are called into question by CGC's specification test as discussed below; see Table 1's columns 4 and 5).

TABLE 1—SUMMARY OF BASELINE AND OB SUBJECTS' ESTIMATED TYPE DISTRIBUTIONS

Type	Apparent from guesses	Econometric from guesses	Econometric from guesses, excluding random	Econometric from guesses, with specification test	Econometric from guesses and search, with specification test
L1	20	43	37	27	29
L2	12	20	20	17	14
L3	3	3	3	1	1
D1	0	5	3	1	0
D2	0	0	0	0	0
Eq.	8	14	13	11	10
Soph.	0	3	2	1	1
Unclassified	45	0	10	30	33

Note: The far-right-hand column includes 17 OB subjects classified by their econometric-from-guesses type estimates.

The hypothesis that $\varepsilon = 1$ is rejected for all but seven of 88 subjects, so the spike is necessary.

The hypothesis that $\lambda = 0$ (payoff-insensitivity) is rejected for 34 subjects.

Thus, payoff-sensitive logit errors significantly improve the fit over a spike-uniform model like CGCB's for only 34/88 = 39% of the subjects.

The lack of significant payoff-sensitivity for most subjects suggests that most of their "errors" are either cognitive or due to misspecification.

The hypothesis $\{\lambda = 0$ and $\varepsilon = 1\}$ is rejected at the 5% level for all but ten of 88 subjects (6 *L1*, 2 *D1*, 1 *Equilibrium*, 1 *Sophisticated*).

Thus, the model does significantly better than a completely random model of guesses for 78/88 = 89% of the subjects.

Specification test

For those 45 subjects whose guesses conformed less closely to one of CGC's types, there is room for doubt about whether CGC's specification omits relevant types and/or overfits by including irrelevant types.

To test for this, CGC conducted a specification test comparing the likelihood of each subject's econometric type estimate with the likelihoods of estimates based on 88 *pseudotypes*, each constructed from one of their subject's guesses in the 16 games.

With regard to overfitting, for a subject's type estimate to be credible it should have higher likelihood than at least as many pseudotypes as it would at random: With 8 types, assuming approximately i.i.d. likelihoods, this suggests it should have higher likelihood than 87/8 ≈ 11 pseudotypes.

Some subjects' type estimates do not pass this test, and so are left unclassified in columns 5 and 6 of CGC's Table 1.

TABLE 1—SUMMARY OF BASELINE AND OB SUBJECTS' ESTIMATED TYPE DISTRIBUTIONS

Type	Apparent from guesses	Econometric from guesses	Econometric from guesses, excluding random	Econometric from guesses, with specification test	Econometric from guesses and search, with specification test
L1	20	43	37	27	29
L2	12	20	20	17	14
L3	3	3	3	1	1
D1	0	5	3	1	0
D2	0	0	0	0	0
Eq.	8	14	13	11	10
Soph.	0	3	2	1	1
Unclassified	45	0	10	30	33

Note: The far-right-hand column includes 17 OB subjects classified by their econometric-from-guesses type estimates.

With regard to omitted types, imagine that CGC had omitted a relevant type, say *L2* for concreteness.

The pseudotypes of CGC’s estimated *L2* subjects would then outperform the non-*L2* types estimated for them, and make approximately the same guesses.

Finding such a *cluster*, CGC diagnosed an omitted type, and studied what its subjects’ guesses had in common to reveal its decision rule.

CGC found five small clusters involving 11 of the 88 subjects, and the subjects in these clusters were also left unclassified in Table 1.

TABLE 1—SUMMARY OF BASELINE AND OB SUBJECTS' ESTIMATED TYPE DISTRIBUTIONS

Type	Apparent from guesses	Econometric from guesses	Econometric from guesses, excluding random	Econometric from guesses, with specification test	Econometric from guesses and search, with specification test
L1	20	43	37	27	29
L2	12	20	20	17	14
L3	3	3	3	1	1
D1	0	5	3	1	0
D2	0	0	0	0	0
Eq.	8	14	13	11	10
Soph.	0	3	2	1	1
Unclassified	45	0	10	30	33

Note: The far-right-hand column includes 17 OB subjects classified by their econometric-from-guesses type estimates.

The paper and its web appendix discuss what these 11 subjects seemed to be doing; most of it appears quite idiosyncratic.

Because a cluster must contain at least two subjects, it is reasonable to anticipate finding more than the five CGC found in a larger sample.

But because any such clusters did not reach the two-subject threshold in CGC's sample of 88, they are probably at most 2% of any larger sample, hence probably not worth modeling.

Taking the specification test into account (as in the right-most column of Table 1 above), econometric estimates of subjects' types are concentrated on *L1*, *L2*, *L3*, and *Equilibrium*, in roughly the same proportions as the subjects whose types are apparent from their guesses.

TABLE 1—SUMMARY OF BASELINE AND OB SUBJECTS' ESTIMATED TYPE DISTRIBUTIONS

Type	Apparent from guesses	Econometric from guesses	Econometric from guesses, excluding random	Econometric from guesses, with specification test	Econometric from guesses and search, with specification test
L1	20	43	37	27	29
L2	12	20	20	17	14
L3	3	3	3	1	1
D1	0	5	3	1	0
D2	0	0	0	0	0
Eq.	8	14	13	11	10
Soph.	0	3	2	1	1
Unclassified	45	0	10	30	33

Note: The far-right-hand column includes 17 OB subjects classified by their econometric-from-guesses type estimates.

Note that unlike the often-suggested interpretation of previous guessing results—that subjects are performing finitely iterated dominance—separating *Lk* from *Dk-1* reveals that *Dk* types don’t exist in any significant numbers, at least in this setting.

Further, CGC’s results for robot/trained subjects, discussed below, suggest that people find doing iterated dominance highly unnatural—as opposed to following *Lk* types that make *k*-rationalizable decisions, and so respect finitely iterated dominance without explicitly performing it.

Sophisticated, which is clearly separated from *Equilbrium* here, as it tends to be when not all subjects play equilibrium strategies, also doesn’t exist.

Econometric puzzles regarding CGC's analysis of decisions

Although CGC's specification test addresses the possibility of bias due to omitting relevant types and/or overfitting by including irrelevant types, it is reasonable to ask if there any way to estimate the distribution of subjects' decision rules without imposing an a priori list of possible types.

However types are determined, they must be general decision rules that are meaningful in any new game.

That is, they cannot just be lists of predicted guesses in CGC's 16 games.

There are at least three reasons for this:

- A worthy competitor to equilibrium must be a general decision rule.
- Allowing completely unrestricted types makes it possible to overfit by defining types like *Miguel* and *Vince* that just happen to do what Miguel and Vince did in the sample
- Because a type's search implications depend not only on what guesses it implies, but why, and types like *Miguel* and *Vince* give us no way to predict what they will do beyond the games we estimated them for.

But the space of possible types is enormous and it has little mathematical structure: Just to avoid ruling out equilibrium, it may have to allow all (even discontinuous) piecewise linear functions of the targets and limits.

Further, conventional clustering analyses rely heavily on Euclidean distance, but without a priori types (whose beliefs imply deviation costs, as required for logit errors) it seems hard to find a credible definition of what it means for subjects' decision patterns to be close.

(For this reason CGC's specification test's analysis of clusters gives more weight to qualitative and structure-dependent patterns of deviation from a reference pattern, such as the tendency, discussed below, of our *Equilibrium* subjects with the clearest fingerprints to deviate much more often in games with mixed targets, and always in the direction of *L3*.)

Finally, it is natural to ask if there is better way to do the specification test.

Questions left unresolved by CGC's analysis of decisions

Some questions regarding subjects' strategic thinking are not resolved by analyzing decisions, but might be resolved by analyzing searches.

Here it is necessary to distinguish CGC's three kinds of treatment.

In the Baseline, subjects played the games with other subjects, looking up both subjects' targets and limits via an interface as explained below.

Open Boxes ("OB") was identical to the Baseline, except that both subjects' targets and limits were continually displayed.

(All the analysis discussed above pooled the data from CGC's Baseline and OB treatments, which did not differ significantly.)

Six different Robot/Trained Subjects ("R/TS") treatments were identical to the Baseline, except subjects played against a "robot" ("the computer") and the computer played according to a pre-specified, announced type, either *L1*, *L2*, *L3*, *D1*, *D2*, or *Equilibrium*; subjects were trained to identify that type's guesses and paid for their payoffs against the computer.

Puzzle A. What are the Baseline "*Equilibrium*" subjects really doing?

Consider the 8 Baseline or OB subjects with near-*Equilibrium* fingerprints:

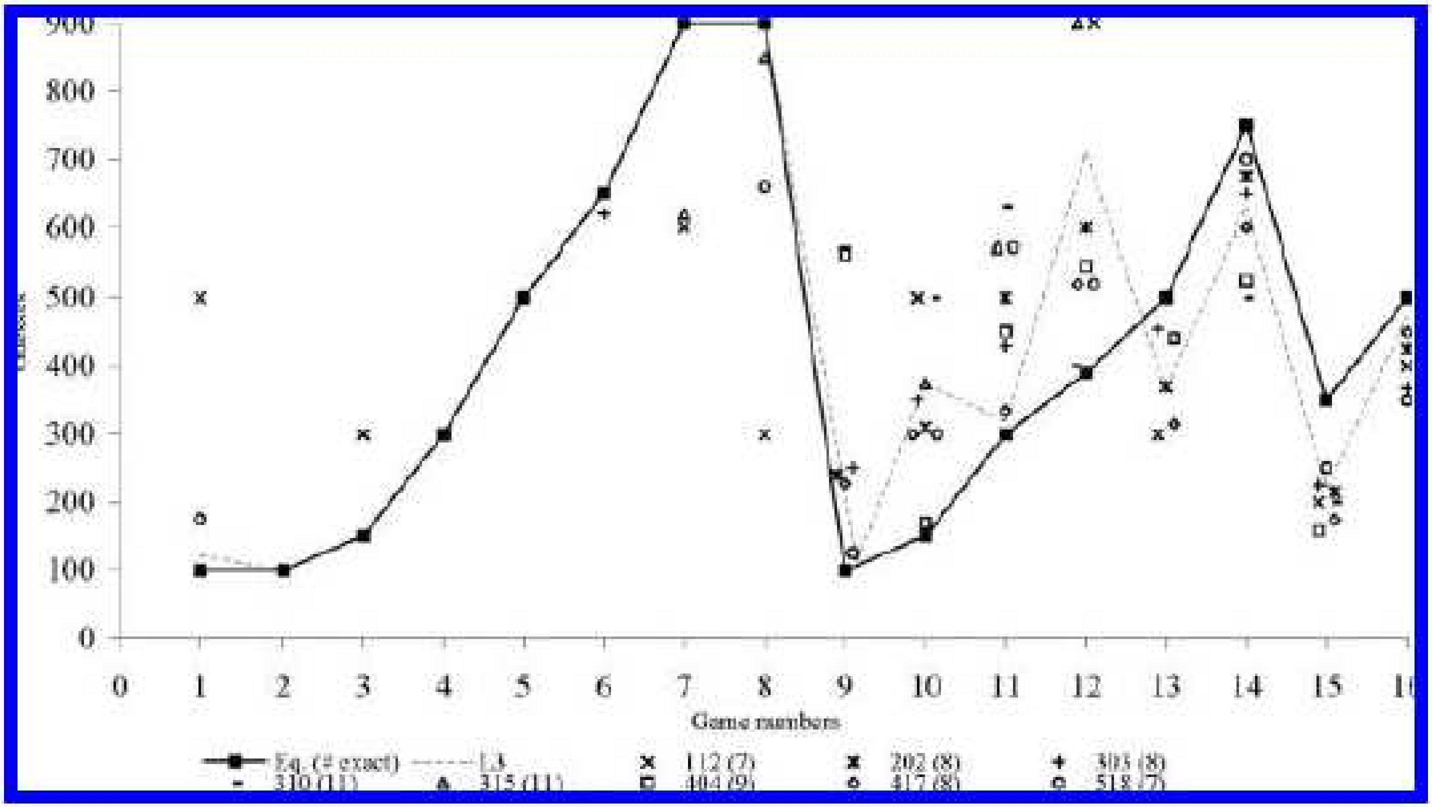


FIGURE 4. "FINGERPRINTS" OF EIGHT APPARENT *EQUILIBRIUM* SUBJECTS

Notes: Only deviations from *Equilibrium*'s guesses are shown. Of these subjects' 128 guesses, 69 (54 percent) were exact *Equilibrium* guesses.

Ordering the games by strategic structure as in CGC's Figure 4, with the eight games with mixed targets (CGC's Table 3, not reproduced here) on the right, shows that those 8 subjects' deviations from equilibrium almost all (50 out of 59, or 85%) occurred in games with mixed targets.

Thus those subjects, whose exact compliance with *Equilibrium* guesses was off the scale by normal standards, are actually following a rule that only mimics *Equilibrium*, and that only in games without mixed targets.

Yet all of the ways we teach people to identify equilibria (best-response dynamics, equilibrium checking, or iterated dominance) work equally well with and without mixed targets: Whatever these subjects were doing, it's something we haven't thought of yet.

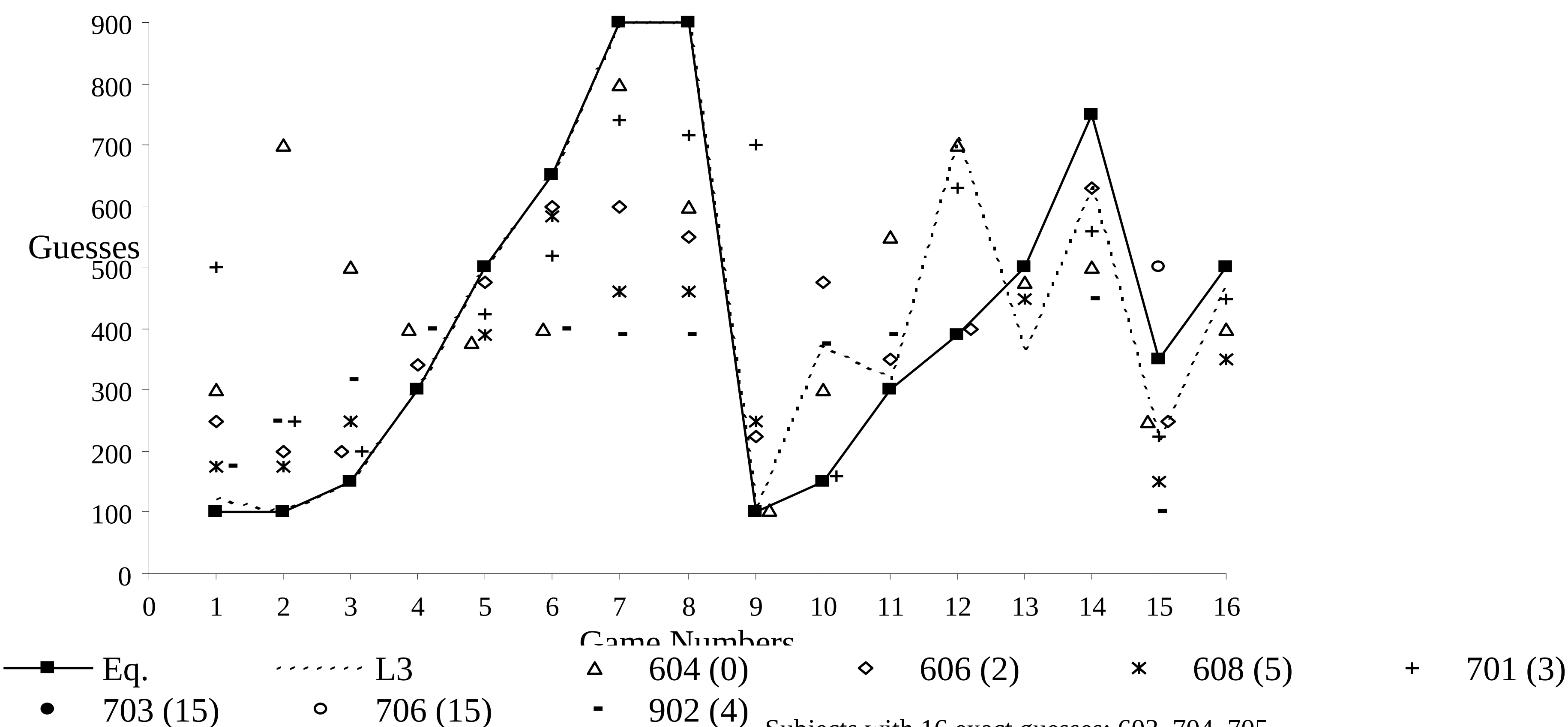
(And their debriefing questionnaires don't tell us what it is.)

Whatever it is, it has some structure: *All* 44 of these subjects' deviations from *Equilibrium* (solid line) when it is separated from *L3* (dotted line) are in the direction of (and sometimes beyond) *L3* guesses.

However, this structure could reflect nothing more than the fact that *Equilibrium* guesses are more extreme than other types' guesses.

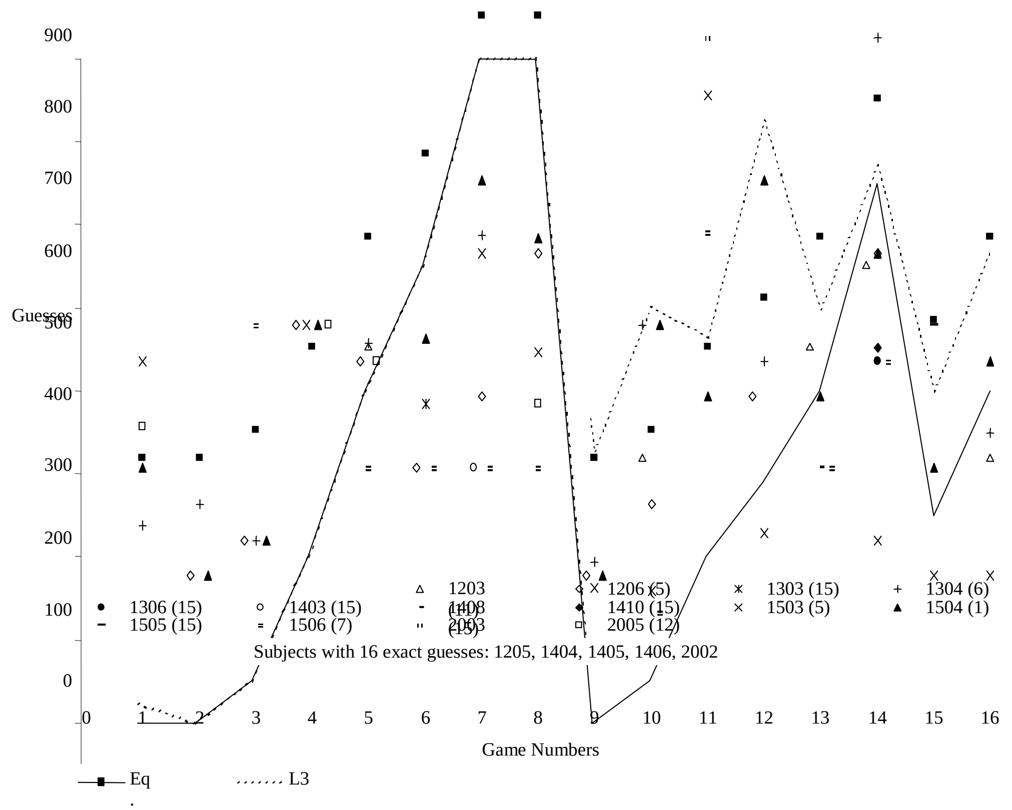
Equilibrium R/TS subjects' compliance is as high with as without mixed targets, so training eliminates whatever the Baseline subjects were doing:

Fingerprints of 10 UCSD *Equilibrium* R/TS Subjects
(only deviations from *Equilibrium*'s guesses are shown)



Subjects with 16 exact guesses: 603, 704, 705

Fingerprints of 18 York *Equilibrium* R/TS Subjects
(only deviations from *Equilibrium*'s guesses are shown)



Puzzle B. Why are *Lk* the only non-*Equilibrium* types that exist?

Recall that a careful analysis of CGC's decision data reveals many subjects of types *L1*, *L2*, *Equilibrium*, or hybrids of *L3* and/or *Equilibrium*, but no other types that do better than a completely random model of guesses for more than one of 88 Baseline/OB subject.

Why do these few rules predominate out of myriads of possible rules?

Why, for instance, aren't there *Dk* subjects, closer to what we teach?

Answering this question may shed some light on bounded rationality.

We suggest possible explanations of both puzzles after discussing CGC's analysis of information search.

CGC's design for studying cognition via information search

In CGC's design for studying cognition via information search, within a publicly announced structure each game was presented via MouseLab, which normally concealed the targets and limits but allowed subjects to look them up as often as desired, one at a time, by clicking on the boxes.

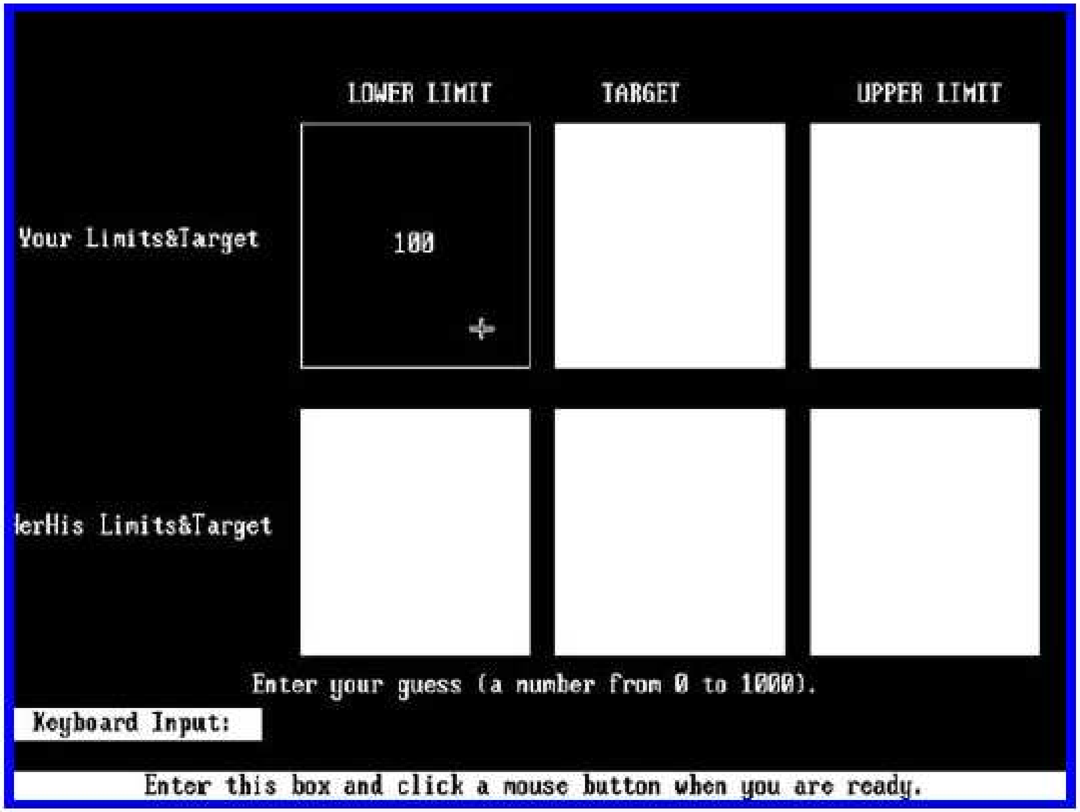


FIGURE 6. SCREEN SHOT OF THE MOUSELAB DISPLAY

CGC's Figure 6. Screen Shot of the MouseLab Display

Details:

CGC used the click option in MouseLab, versus CJ’s use of the rollover option.

Thus, opening and closing boxes both required conscious decisions.

Subjects were not allowed to write, and the data strongly suggest that subjects did not memorize the targets and limits.

With search costs as low as subjects’ searches make them seem, free access made the entire structure effectively public knowledge, so the results can be used to test theories of behavior in complete-information versions of the games.

The design also maintains control over subjects’ motives for search by making information from previous plays irrelevant to current payoffs.

From the point of view of studying cognition via search, CGC's normal-form design combines the strengths of CJ's extensive-form design and CGCB's matrix-game design.

CJ's extensive-form design allows subjects to search for a small number of hidden payoff parameters (pies in alternating-offers bargaining) within a simple, publicly announced structure.

However, it also makes subjects' search patterns essentially one-dimensional, and so less informative than they could be.

CGC’s design maintains the simplicity of CJ’s design, allowing subjects to focus on predicting others’ decisions without getting lost in the details of the structure.

Unlike CJ’s design but like CGCB’s, CGC’s design makes search higher-dimensional, hence more informative.

Like CGCB’s design, CGC’s design also independently separates types’ implications for search and decisions, revealing relationships between them.

But unlike CGCB’s design, CGC’s makes types’ search implications almost independent of the game, an important convenience in analysis.

Search data for representative R/TS and Baseline subjects

We start by comparing search data for representative R/TS and Baseline subjects whose guesses conform closely to their assigned or estimated type with the implications of CGC's theory of cognition and search.

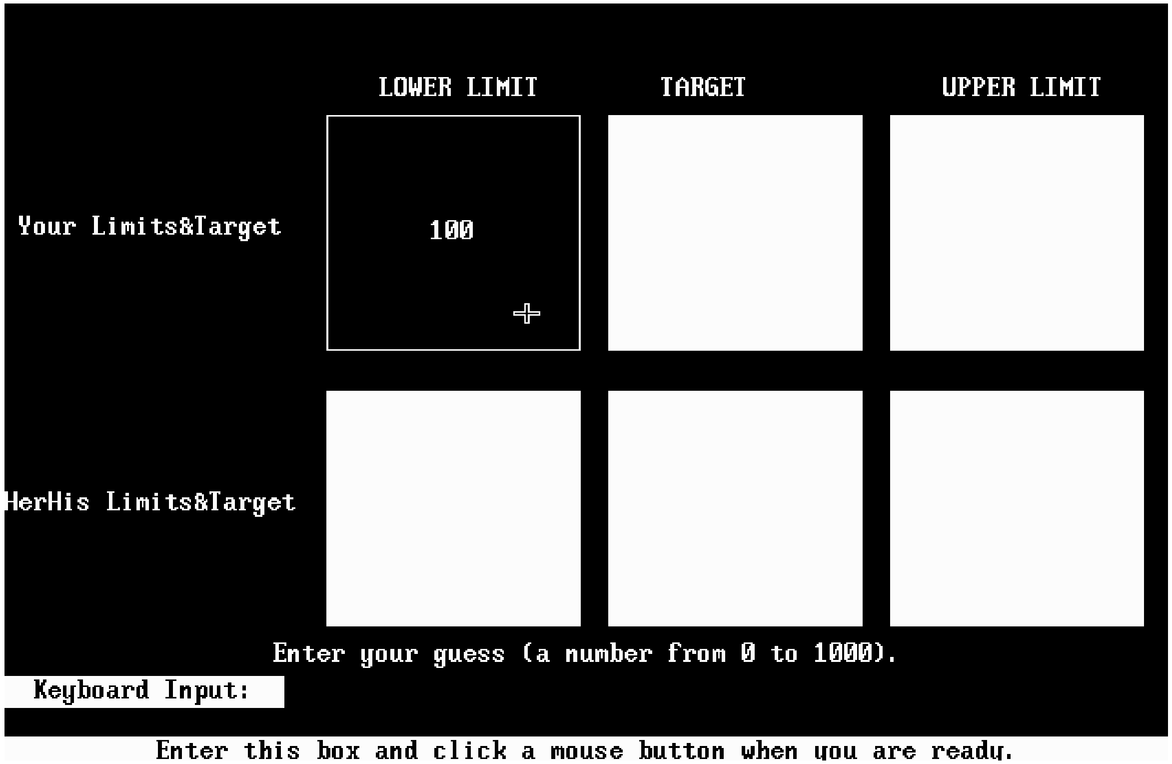
Eyeballing compliance with the types' search implications will suggest that there is some usable structure in the data, and provide some hints about how to model it.

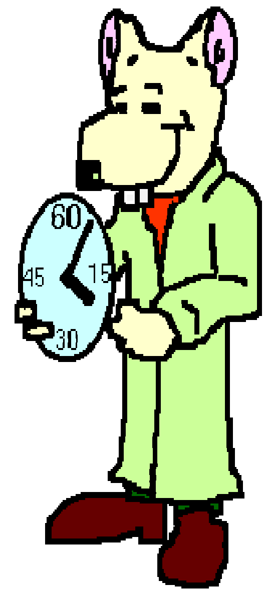
We will then explain CGC's (and CGCB's) theory of cognition and information search, show how the search implications were derived, and show how to use them to model subjects' searches econometrically.

(Because CGC's theory is close to CGCB's, it was almost completely specified before these data were generated).

But first….

Speak rodent like a native in one easy lesson!





	a	***p***	***b***
You (*i*)	1	2	3
S/he (*j*)	4	5	6

MouseLab box numbers

Selected R/TS Subjects' Information Searches and Assigned Types' Search Implications

MouseLab box

	a	*p*	*b*
You (*i*)	1	2	3
S/he (*j*)	4	5	6

Types' Search Implications

L1	{[4,6],2}
L2	{([1,3],5),4,6,2}
L3	{([4,6],2),1,3,5}
D1	{(4,[5,1], (6,[5,3]),2}
D2	{(1,[2,4]),(3,[2,6]),(4,[5,1],(6,[5,3]),5,2
Eq	{[2,5],4} if pr. tar.<1,{[2,5],6} if > 1

Subject	**904**	**1716**	**1807**	**1607**	**1811**	**2008**	**1001**	**1412**	**805**	**1601**	**804**	**1110**	**1202**	**704**	**1205**	**1408**	**2002**
Type(#rt.	L1 (16)	L1 (16)	L1 (16)	L2	L2	L2	L3	L3	D1	D1	D1 (3)	D2	D2	Eq	Eq	Eq	Eq (16)
Alt.(#rt.)											L2						
Est. stvle	late	often	earlv	often	earlv				Earlv								
Game																	
1	123456 4623	146462 134646 23	462513	135462 1313	134446 5213*4 6	111313 131313 5423	462135 21364* 246231 52	146231 564623 1	154356 423213 2642	254514 36231	154346 5213	135464 2646*1 313	246466 135464 641321 342462 422646 124625 5*1224 654646	123456 363256 565365 626365 652651 452262 6526	123456 424652 562525 6352*4 65	123123 456445 632132 11	142536 125365 253616 361454 613451 213452 63
2	123456 4231	462462 13	462132 25	135461 354621 3	134653 125642 313562 52	131313 566622 333	462135 642562 223146 2562*6 2	462462 546231 546231	514535 615364 23	514653 6213	515135 365462 3	135134 642163 451463 211136 414262 135362 *14654 6	123645 132462 426262 241356 462*13 524242 466135 6462	123456 525123 652625 635256 262365 456	123456 244565 565263 212554 146662 654251 44526* 31	123456 456123 643524 1	143625 361425 142523 625656 3

Notes: The subjects' frequencies of making their assigned types' (and when relevant, alternate types') exact guesses are in parentheses after the assigned type. A * in a subject's look-up sequence means that the subject entered a guess there without immediately confirming it.

Selected Baseline Subjects' Information Searches and Estimated Types' Search Implications

	MouseLab box		
	a	*p*	*b*
You (*i*)	1	2	3
S/he (*j*)	4	5	6

Types' Search Implications

L1	{[4,6],2}
L2	{([1,3],5),4,6,2}
L3	{([4,6],2),1,3,5}
D1	{(4,[5,1], (6,[5,3]),2}
D2	{(1,[2,4]),(3,[2,6]),(4,[5,1],(6,[5,3]),5,2
Eq	{[2,5],4} if pr. tar.<1,{[2,5],6} if > 1

Subject	**101**	**118**	**413**	**108**	**206**	**309**	**405**	**210**	**302**	**318**	**417**	**404**	**202**	**310**	**315**
Type(#rt.	L1 (15)	L1 (15)	L1 (14)	L2	L2	L2	L2	L3 (9)	L3 (7)	L1 (7)	Eq (8)	Eq (9)	Eq (8)	Eq	Eq
Alt.(#rt.)								Eq (9)	Eq (7)	D1 (5)	L3 (7)	L2 (6)	D2 (7)		
Alt.(#rt.)								D2 (8)			L2 (5)		L3 (7)		
Est. style	early/la	early	late	early	early	early/la	early	early	Early	early	early	early	early	early/la	early
Game															
1	146246 213	246134 626241 32*135	123456 545612 3463*	135642	533146 213	1352	144652 313312 546232 12512	123456 123456 213456 254213 654	221135 465645 213213 45456* 541	132456 465252 13242* 1462	252531 464656 446531 641252 462121 3	462135 464655 645515 21354* 135462 426256 356234 131354 645	123456 254613 621342 *525	123126 544121 565421 254362 *21545 4*	213465 624163 564121 325466
2	46213	246262 2131	123564 62213*	135642 3	531462 31	135263 1526*2 *3	132456 253156 456545 463123 156562 62	123456 465562 231654 456*2	213546 566213 545463 21*266 54123	132465 132*46 2	255236 62*365 243563	462461 352524 261315 463562	123456 445613 255462 513565 23	123546 216326 231456 *62	134652 124653 656121 3
3	462*46	246242 466413 *426	264231	135642 53	535164 2231	135263	312456 5231*1 236545 5233** 513	123455 645612 3	265413 232145 563214 563214 523*65 4123	134652 1323*4	521363 641526 5263*6 52	462135 215634 *52	123456 123562 3	123655 463213	132465 544163 *3625

These data suggest the following conclusions:

(i) Search is so heterogeneous and noisy that we should study it at the individual subject level.

(ii) There is little difference between the look-up sequences of R/TS and Baseline subjects of a given type (assigned type for R/TS, apparent type for Baseline), except that the R/TS look-up sequences are usually shorter than the Baseline ones. (Perhaps the small difference is unsurprising, because R/TS subjects were not trained in search strategies.)

(iii) A subject's type's predicted look-up sequence is unusually dense in his searches, at least for types *L1* and *L2*, and one can quickly learn to read the algorithms many subjects are using directly from the data.

(iv) For some subjects search is an important check on decisions; for example, Baseline subject 309, with 16 exact *L2* guesses, missed some of *L2*'s relevant look-ups in the first few games, avoiding deviations from *L2* only by luck. (S/he had a Eureka! moment between games 5 and 6, and from then on complied perfectly.) This recalls CJ's finding that in their alternating-offers bargaining games, 10% of the subjects *never* looked at the last-round pie and 19% never looked at the second-round pie.

How does cognition show up in information search?

In studying cognition via information search, CJ followed the tradition in the psychology literature, giving roughly equal weight to look-up durations and to the numbers of look-ups of each pie (“acquisitions”) and the transitions between pies.

Gabaix, Laibson, Moloche, and Weinberg, “Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model,” *AER* 2006, focused on acquisitions and considered aspects of look-up order too.

Rubinstein *EJ* 2007, which considers some matrix games, considered only durations.

These analyses were mostly conducted at a high level of aggregation, both across subjects and over time.

By contrast, CGC, following CGCB, took it as a given that cognition is sufficiently heterogeneous and search sufficiently noisy that they are best studied at the individual level.

CGC and CGCB also assumed that which look-ups subjects make, in which order, are at least as revealing as look-up durations or acquisition frequencies.

(CGC and CGCB made no claim that durations are irrelevant, just that they don't deserve the top priority they have been given.

CGCB present some results on durations, "gaze times" in their Table IV.)

CGC's views were shaped by simple-minded theories of cognition, CJ's R/TS searches, and CGCB's *Equilibrium* Trained Subjects' searches.

Thinking types as models of cognition and search

CGC’s (and CGCB’s) models of cognition, search, and decisions are based on a procedural view of decision-making, in which a subject’s type determines his search, and type and search then determine his decision.

Each type is naturally associated with algorithms that process payoff information into decisions. (As noted above, because a type’s search implications depend not only on what decisions it specifies, but why, something like a types-based model seems necessary here.)

The analysis uses the algorithms as models of cognition, deriving a type’s search implications under simple assumptions about how cognition determines search.

The types then provide a basis for the enormous space of possible decision and search sequences, imposing enough structure to describe subjects’ behavior in a comprehensible way, and to make it meaningful to ask how subjects’ decisions and searches are related.

How Does Cognition Determine Search?

Without further assumptions, nothing precludes a subject’s scanning and memorizing the information and then “going into his brain” to figure out what to do, in which case his searches will reveal nothing about cognition.

But inspecting the sample of actual searches above suggests that there are strong regularities in search behavior, and that subjects’ searches might therefore contain a lot of information about cognition.

The goal in analyzing search is to add enough assumptions to make it possible to extract the signal from the noise in subjects’ look-up sequences; but not so many that they distort the meaning of the signal.

CGC’s (and CGCB’s) assumptions are conservative, resting on types’ minimal search implications and adding as little structure as possible.

Types’ Search Implications

CGC derived types’ minimal search implications from their *ideal guesses*, those they would make if they had no limits. (With automatic rounding of guesses and quasiconcave payoffs, ideal guesses are all that subjects need to know, and all that matters for minimal search implications.)

Evaluating a formula for a type’s ideal guess requires a series of *operations*, some of which are *basic* in that they logically precede any other operation.

For example, $[a^j+b^j]$ (averaging the partner’s limits) is the only basic operation for *L1*'s ideal guess, $p^i[a^j+b^j]/2$.

CGC derived types’ search implications assuming that subjects perform basic operations one at a time via adjacent look-ups, remember their results, and otherwise rely on repeated look-ups rather than memory.

These empirically-based assumptions seem to yield a reasonably accurate model of most subjects’ search behavior.

The left side of Table 4 on the next slide lists the formulas for types' ideal guesses in CGC's games.

The right side of Table 4 lists types' minimal search implications, derived as just explained: first in terms of our notation, then in terms of the box numbers in which MouseLab records the data.

Basic operations are represented by adjacent look-up pairs that can appear in either order, but cannot be separated by other look-ups.

Such pairs are grouped within square brackets, as in $\{[a^j, b^j], p^i\}$ for *L1*.

Other operations can appear in any order and their look-ups can be separated.

Such operations are represented by look-ups grouped within curly brackets or parentheses.

A type's operations are listed in the order that seems most natural, if there is one; but this is not a requirement of the theory.

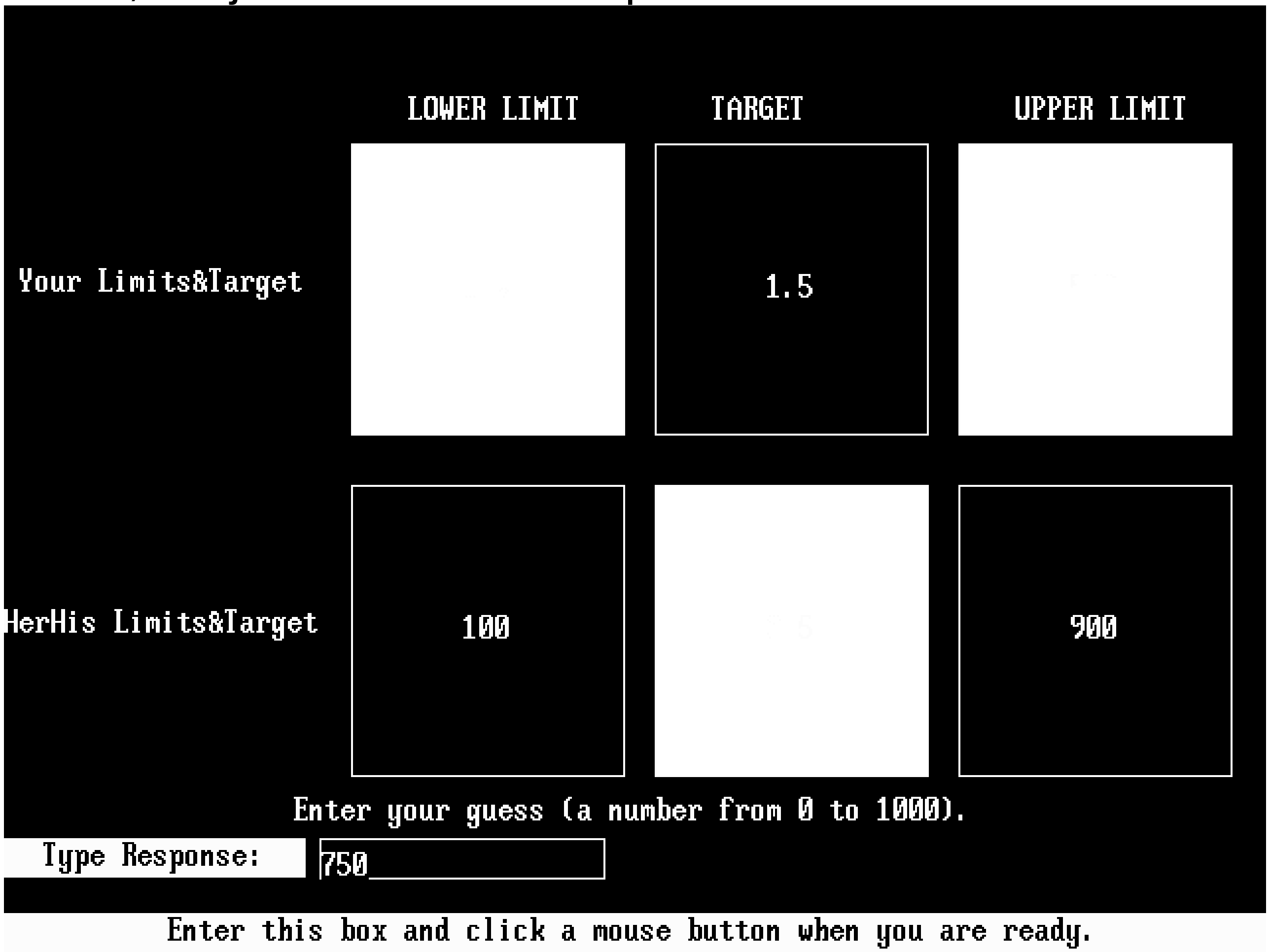
Type	Ideal guess	Relevant look-ups
L1	$p^i\,[a^j+b^j]/2$	$\{[a^j,b^j],p^i\} \equiv \{[4, 6], 2\}$
L2	$p^iR(a^j,b^j;\ p^j[a^i+b^i]/2)$	$\{([a^i,b^i],p^j),a^j,b^j,p^i\} \equiv \{([1, 3], 5), 4, 6, 2\}$
L3	$p^iR(a^j,b^j;\ p^jR(a^i,b^i;\ p^i[a^j+b^j]/2))$	$\{([a^j,b^j],p^i),a^i,b^i,p^j\} \equiv \{([4, 6], 2), 1, 3, 5\}$
D1	$p^i(\max\{a^j,p^ja^i\} + \min\{p^jb^i,b^j\})/2$	$\{(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^i\} \equiv \{(4,[5,1]),(6,[5,3]),2\}$
D2	$p^i[\max\{\max\{a^j,p^ja^i\},p^j\max\{a^i,p^ia^j\}\} + \min\{p^j\min\{p^ib^j,b^i\},\min\{p^jb^i,b^j\}\}]/2$	$\{(a^i,[p^i,a^j]),(b^i,[p^i,b^j]),(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^j,p^i\} \equiv \{(1,[2,4]),(3,[2,6]),(4,[5,1]),(6,[5,3]),5,2\}$
Eq.	p^ia^j if $p^ip^j < 1$ or p^ib^j if $p^ip^j > 1$	$\{[p^i,p^j],a^j\} \equiv \{[2, 5], 4\}$ if $p^ip^j < 1$ or $\{[p^i,p^j],b^j\} \equiv \{[2, 5], 6\}$ if $p^ip^j > 1$
Soph.	[no closed-form expression, but CGC took its search implications to be the same as *D2*'s]	$\{(a^i,[p^i,a^j]),(b^i,[p^i,b^j]),(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^j,p^i\} \equiv \{(1,[2,4]),(3,[2,6]),(4,[5,1]),(6,[5,3]),5,2\}$

CGC's Table 4. Types' Ideal Guesses and Relevant Look-ups

(*p* is a target; *a* (*b*) is a lower (upper) limit; *i* and *j* are the player and his partner; and $R(\cdot)$ is the automatic adjustment function.)

L1's search implications

(Unlike in this picture, subjects could never open more than one box at a time.)

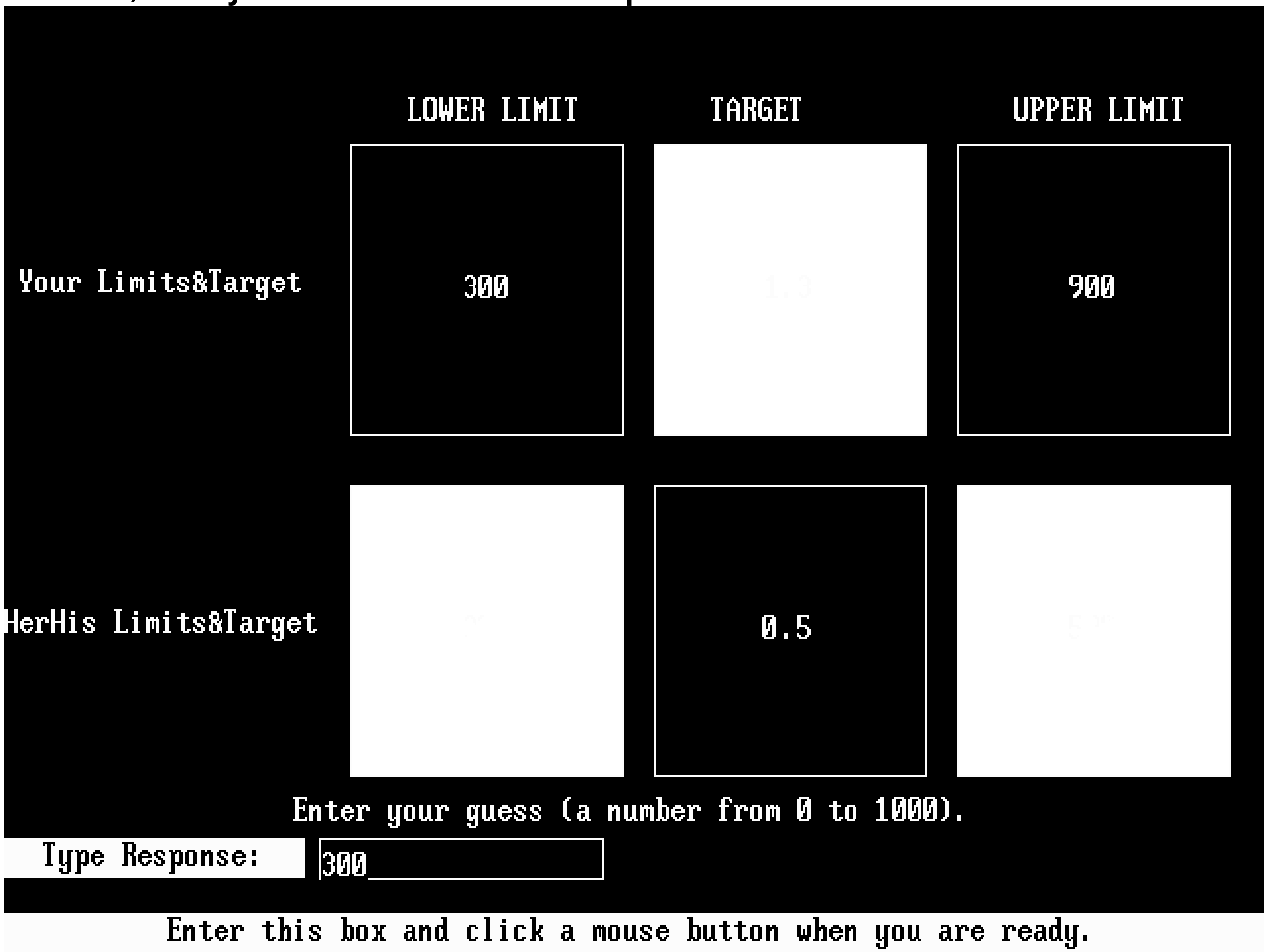


L1's ideal guess: $p^i[a^j+b^j]/2 = 750$. *L1*'s search implications: $\{[a^j, b^j], p^i\} \equiv \{[4, 6], 2\}$.

(*L1* does not need to look up its own limits because it can enter its ideal guess and rely on automatic adjustment to ensure that its adjusted guess is optimal. Thus this design even separates *L1* from a *Solipsistic* type that only looks up its own parameters.)

L2’s search implications: first step

(Unlike in this picture, subjects could never open more than one box at a time.)



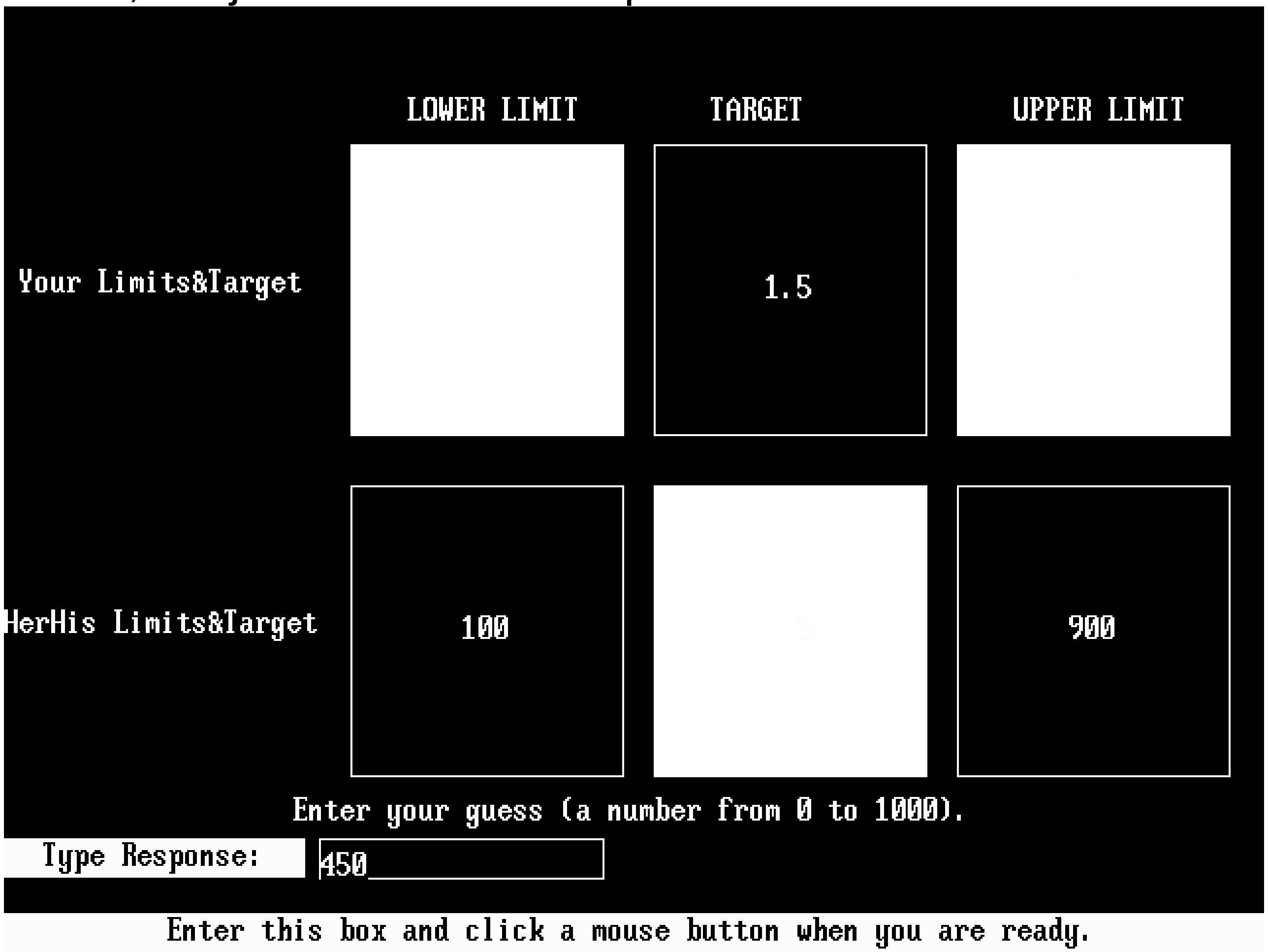
L2’s model of its partner’s *L1* guess: $p^j[a^j+b^j]/2 = 300$.

Search implications: $\{[a^i,b^i],p^j\} \equiv \{[1,3],5\}$.

(*L2* needs to look up its own limits only to predict its partner’s *L1* guess; like *L1* it can enter its ideal guess and rely on automatic adjustment to ensure its adjusted guess is optimal.)

L2’s search implications: second step

(Unlike in this picture, subjects could never open more than one box at a time.)



L2’s ideal guess: $p^i R(a^j,b^j;\ p^j[a^i+b^i]/2) = 450$.

L2’s search implications: $\{([a^i,b^i],p^j),a^j,b^j,p^i\} \equiv \{([1, 3], 5), 4, 6, 2\}$.

(*L2* needs to look up its partner’s limits $a^j = 4$ and $b^j = 6$ to predict its partner’s *L1 adjusted* guess.)

Aside on types' search implications

L1, *L2*, *L3*, *D1*, *D2* search implications are easy to derive from the formulas in Table 4.

Note that although most theorists instinctively identify *Lk* with *Dk*-1, etc., they are cognitively very different:

Lk starts with a naïve prior over the other's decisions and iterates the best-response mapping; *Dk*-1 starts with reasoning based on iterated knowledge of rationality and closes the process with a naïve prior.

This difference shows up clearly in their search implications in Table 4:

$\{([a^i,b^i],p^j),a^j,b^j,p^i\} \equiv \{([1, 3], 5), 4, 6, 2\}$ for *L2*

versus

$\{(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^i\} \equiv \{(4,[5,1]),(6,[5,3]),2\}$ for *D1*.

Equilibrium can use any workable method to find its ideal guess; we allow any method, and seek the one with minimal search requirements.

Equilibrium-checking (conjecturing guesses and checking them for consistency with equilibrium) is less demanding than other methods, but requires more luck than almost all of our subjects appeared to have.

Accordingly, we allow *Equilibrium* to use both targets to determine whether the equilibrium is High or Low, and then to enter its own target times its partner's lower (upper) limit when the product of targets is < (>) 1, which CGC's Observation 1 shows ensures its adjusted guess is in equilibrium.

This has the same search requirements as equilibrium-checking except that it requires the targets to be adjacent; and thereby avoids the need for luck.

(Unlike in CGCB's and CJ's designs, *Equilibrium*'s search implications are just as simple as *L1*'s, and simpler than other boundedly rational types'!)

(End of aside)

Searchmetrics

CGC’s econometric analysis of guesses and search extends CGC’s (and CGCB’s) maximum likelihood error-rate models of decisions to explain search compliance as well as decisions, treating search as just another kind of decision as much as possible.

The main econometric problem is extracting signals from subjects' highly idiosyncratic, noisy look-up sequences, without a well-tested model that implies strong restrictions on how cognition drives search.

Among other things, subjects vary in the location of look-ups relevant to their types in their sequences.

CGC filter this out via subject-specific nuisance parameters called style (“early” or “late”), assumed constant across games for each subject.

(58 of 71 Baseline subjects’ estimated styles are early, 10 are late, and 3 are tied.)

CGC summarize a subject's compliance with a type's search implications in a game by the density of the type's look-up sequence in the relevant part (as determined by estimated style) of the subject's look-up sequence.

If, for example, style is early, a subject's search compliance for a given type is defined by starting at the beginning of his look-up sequence and continuing until the type's relevant sequence (Table 4) is first completed.

Compliance is then the length of the relevant sequence divided by the length of the sequence that first completed it.

This definition filters out irrelevant look-ups (except if they separate the adjacent look-ups required for a basic operation) in a simple way, while making compliance meaningfully comparable across games and styles.

CGC assume that a subject's type and style determine his search and guess in a given game, each with error.

They further assume that, given type and style, errors in search and guesses are independent of each other and across games.

(This strong but useful simplifying assumption makes the log-likelihood separable across guesses and search, avoiding complications in CGCB.)

To avoid stronger distributional assumptions CGC discretized compliance into three categories: $C_H \equiv [0.67,1.00]$, $C_M \equiv [0.33,0.67]$, and $C_L \equiv [0,0.33]$.

Subject *i*'s guesses-and-search log-likelihood is:

$$\sum_c \left[m_c^{isk} \ln(\zeta_c) + (m_c^{isk} - n_c^{isk}) \ln(1-\varepsilon) + n^{isk} \ln(\varepsilon) + \sum_{g \in N_c^{isk}} \ln d_g^k (R_g^i(x_g^i), \lambda) \right] \equiv$$

$$(G-n^{ik})\ln(G-n^{ik}) + n^{ik}\ln(n^{ik}) + \sum_{g \in N^{ik}} \ln d_g^k (R_g^i(x_g^i), \lambda) - G\ln G + \sum_c \left[m_c^{isk} \ln m_c^{isk} \right] - 2G\ln G,$$

where m_c^{isk} is the number of games for which subject *i* has type-*k* style-*s* compliance *c*.

The search term is convex in the m_c^{isk}, and therefore favors types for which compliance varies less across games, because such types "explain" search behavior better. See CGCB, Section 4.D.

The maximum-likelihood estimates of ε and ζ_c, given *k* and *s*, are n^{ik}/G and m_c^{isk}/G, the sample frequencies with which subject *i*'s adjusted guesses are non-exact for that *k* and *i* has compliance *c* for that *k* and *s*.

The maximum likelihood estimate of *λ* is the standard logit precision.

The maximum likelihood estimate of subject *i*'s type *k* maximizes the above log-likelihood over *k* and *s*, given the estimated ε and *λ*.

Note that the model favors such types without regard to whether compliance is high or low.

This seems appropriate because compliance is neither meaningfully comparable across types (as opposed to across games and styles); nor is it guaranteed to be high for the "true" type (which could be cognitively very difficult).

But it means that CGC need to rule out estimates where a type wins simply because its compliance is very low in all games.

Most guesses-and-search type estimates, especially those for subjects whose guess fingerprints were clear, reaffirm guesses-only estimates.

Thus, overall, incorporating search into the econometric analysis confirms our conclusions, including the absence of significant numbers of subjects of types other than *L1*, *L2*, *Equilibrium*, or hybrids of *L3* or *Equilibrium*.

Incorporating search does refine and sharpens our conclusions in some ways; and a few subjects' type estimates change (Table 1, 7A, and 7B).

TABLE 1—SUMMARY OF BASELINE AND OB SUBJECTS' ESTIMATED TYPE DISTRIBUTIONS

Type	Apparent from guesses	Econometric from guesses	Econometric from guesses, excluding random	Econometric from guesses, with specification test	Econometric from guesses and search, with specification test
L1	20	43	37	27	29
L2	12	20	20	17	14
L3	3	3	3	1	1
D1	0	5	3	1	0
D2	0	0	0	0	0
Eq.	8	14	13	11	10
Soph.	0	3	2	1	1
Unclassified	45	0	10	30	33

Note: The far-right-hand column includes 17 OB subjects classified by their econometric-from-guesses type estimates.

For some subjects the guesses-and-search estimate resolves a tension between guesses-only and search-only estimates in favor of a type other than the guesses-only estimate.

The search part of the likelihood has weight only about 1/6 of the guesses part, because our theory of search makes much less precise predictions than our theory of guesses—a necessary evil, given the noisiness and idiosyncrasy of search behavior.

For other subjects the guesses-only type estimate has 0 search compliance in 8 or more games, and so CGC rule it out a priori.

For example, Baseline 415, with apparent type *L1* with 9 exact guesses, had 0 *L1* search compliance in 9 of the 16 games because s/he had no adjacent $[a^j, b^j]$ pairs as required for *L1*.

However, her/his sequences were unusually rich in (a^j, p^i, b^j) and (b^j, p^i, a^j) triples, in those orders.

Because the sequences were not rich in such triples with other superscripts, we conclude that 415 was a true *L1* who was more comfortable with several numbers in working memory than our characterization assumes, or than our other subjects were comfortable with.

But because this violated our assumptions on search, this subject was "officially" estimated to be a *D1*.

(This is why we downplay the official estimate above.)

Many subjects' types can be reliably identified from search alone (Table 7A):

Table 7A—Subject-by-Subject Guesses-Only and Search-Only Econometric Type Estimates and Results of Specification Test

		Guesses only				Search only			
ID	Dom.	ln L	k	Exact	λ	ln L	k	ξ_H	ξ_M
513	0	0.00	$L1$	16	—	—	—	—	—
118	0	−9.62	$L1$	15	1.85	−7.41	$L1_e$	0.88	0.06
101	1	−10.27	$L1$	15	0.55	−9.94	$L1_e^{\ddagger}$	0.69	0.31
104	0	−16.63	$L1$	14	2.20*	−3.74	$L1_e$	0.00	0.94
413	0	−17.81	$L1$	14	0.88	−6.03	$L1_j$	0.13	0.88
207	0	−17.96	$L1$	14	0.42	0.00	$L1_e$	1.00	0.00
216	1	−25.41	$L1$	13	1.06	−11.25	$L3_e$	0.75	0.19
402	0	−30.93	$L1$	12	5.65*	−9.00	$L1_e$	0.00	0.75
418	0	−42.23	$L1$	10	21.22**	−7.41	$L2_e$	0.88	0.06
301	1	−45.84	$L1^{D}$	10	0.00	−3.74	$L1_e$	0.06	0.94
508	0	−46.19	$L1^{D}$	10	2.05	—	—	—	—
308	3	−47.34	$L1$	10	0.00	−9.63	$L3_e$	0.81	0.13
102	4	−47.63	$L1$	10	0.00	−9.63	$L2_e$	0.81	0.06
415	1	−53.64	$L1$	9	0.88	−16.38	$D1_e$	0.31	0.50
504	1	−56.97	$L1$	8	1.68**	—	—	—	—
208	6	−61.62	$L1$	8	0.00	−3.74	$L1_j$	0.06	0.94
318	0	−62.61	$L1$	7	3.18*	−3.74	$L1_e^{\ddagger}$	0.00	0.94
512	0	−63.33	$L1$	7	1.56	—	—	—	—
502	1	−64.55	$L1$	7	1.01	—	—	—	—
516	1	−64.93	$L1^{C}$	7	1.10*	—	—	—	—
409	0	−73.59	$L1^{E}$	4	9.90**	−10.59	$L1_j$	0.00	0.38
106	0	−75.82	$L1$	5	1.19*	−7.72	Eq_e	0.00	0.19
305	3	−79.89	$L1$	5	0.37	−6.03	$L1_e$	0.88	0.13
411	1	−80.58	$L1$	4	1.45**	0.00	$L3_e$	1.00	0.00
509	1	−81.81	$L1$	4	0.86	—	—	—	—
203	4	−83.90	$L1$	4	0.00	−9.94	Eq_e	0.00	0.31
505	4	−84.13	$L1$	4	0.43	—	—	—	—
317	3	−86.58	$L1$	3	0.92*	−3.74	$L1_e$	0.94	0.06
416	1	−86.74	$L1^{\dagger}$	1	4.48**	−3.74	$L1_e^{\ddagger}$	0.00	0.94
217	3	−87.12	$L1$	3	0.68	−10.59	$L1_e$	0.00	0.38
219	3	−87.32	$L1^{+}$	3	0.89*	−7.72	$L1_e$	0.00	0.81
501	1	−87.93	$L1^{\dagger}$	0	4.38**	—	—	—	—
410	3	−89.18	$L1$	2	1.53**	−7.72	$L1_{el}^{\ddagger}$	0.00	0.19
510	5	−89.60	$L1$	3	0.00	—	—	—	—
420	2	−89.68	$L1^{+}$	2	1.25**	−3.74	Eq_j	0.00	0.06
408	2	−89.71	$L1^{+}$	2	1.09*	−6.03	$L1_e$	0.00	0.88
201	3	−90.26	$L1^{+}$	2	1.21**	−3.74	$L1_e^{\ddagger}$	0.00	0.94
105	2	−90.58	$L1^{+}$	2	1.29**	−9.00	Eq_e	0.25	0.75
103	3	−90.61	$L1^{+}$	2	1.12*	−6.03	$L1_e$	0.00	0.13
213	2	−95.57	$L1^{\dagger+}$	0	1.19*	−3.74	$L2_e$	0.94	0.00
515	4	−95.68	$L1^{\dagger+}$	1	0.60	—	—	—	—
113	5	−96.61	$L1^{\dagger+}$	1	0.07	−9.63	$L3_{el}^{\ddagger}$	0.81	0.06
109	8	−97.31	$L1^{\dagger+}$	1	0.00	—	—	—	—
309	0	0.00	$L2$	16	—	−9.94	$L2_{el}^{\ddagger}$	0.69	0.00
405	0	0.00	$L2$	16	—	−13.30	$L3_e$	0.69	0.13
206	0	−10.07	$L2$	15	0.79	−7.41	$L2_e$	0.88	0.06
209	0	−25.51	$L2$	13	0.96	−9.00	$L1_e$	0.00	0.75
108	0	−25.88	$L2$	13	0.45	0.00	$L2_e^{\ddagger}$	1.00	0.00
214	2	−35.30	$L2$	11	2.73**	−3.74	$L1_e$	0.00	0.94
307	1	−38.88	$L2$	11	1.04*	−7.72	Eq_e	0.00	0.19
218	0	−40.54	$L2$	11	0.60	−7.72	$L1_e$	0.00	0.81
422	2	−55.79	$L2$	9	0.22	0.00	$L1_e$	0.00	1.00
316	1	−58.43	$L2$	8	0.73	−10.97	$Eq_e^{\ddagger}$	0.00	0.44
407	0	−60.98	$L2^{C}$	8	0.44	−6.03	$L2_e^{\ddagger}$	0.88	0.13
306	2	−68.48	$L2$	7	0.18	−3.74	$L1_j$	0.00	0.06
412	0	−69.43	$L2$	6	1.05**	0.00	$L2_e^{\ddagger}$	1.00	0.00

Table 7A—Continued

		Guesses only				Search only			
ID	Dom.	ln L	k	Exact	λ	ln L	k	ξ_H	ξ_M
205	0	−72.81	$L2$	6	0.01	0.00	$L1_e$	0.00	1.00
220	1	−72.96	$L2$	6	0.32	0.00	$L1_e$	0.00	1.00
403	0	−73.60	$L2$	6	0.50	−6.03	$Eq_j^{\ddagger}$	0.00	0.13
517	0	−73.70	$L2$	5	0.98**	—	—	—	—
503	3	−88.21	$L2^{+}$	3	0.00	—	—	—	—
414	4	−89.00	$L2$	2	0.78*	−7.72	$L1_e$	0.00	0.19
110	3	−92.51	$L2^{+}$	2	0.00	−9.00	$L1_j$	0.00	0.75
210	0	−51.13	$L3^{B}$	9	0.92*	−10.59	$L1_e$	0.00	0.38
302	0	−61.46	$L3^{B}$	7	1.11**	−6.03	Eq_e	0.00	0.13
507	0	−63.23	$L3$	7	0.94**	—	—	—	—
313	0	−79.12	$D1^{E}$	3	2.68**	−6.03	$L1_e^{\ddagger}$	0.00	0.88
312	0	−80.45	$D1^{\dagger}$	1	5.85**	−3.74	$L2_e^{\ddagger}$	0.94	0.06
204	2	−84.86	$D1^{E}$	3	1.22**	0.00	$L1_e^{\ddagger}$	0.00	1.00
115	1	−86.10	$D1$	2	1.74**	−9.94	Eq_e	0.00	0.31
401	2	−91.99	$D1^{\dagger+}$	0	1.58**	−6.03	Eq_j	0.00	0.13
310	0	−41.69	Eq	11	0.00	−9.94	$L1_j$	0.00	0.31
315	0	−41.80	Eq	11	0.00	0.00	$L3_e^{\ddagger}$	1.00	0.00
404	1	−54.69	Eq	9	0.03	−9.00	$Eq_e^{\ddagger}$	0.00	0.75
303	0	−59.93	Eq	8	0.41	−3.74	$Eq_e^{\ddagger}$	0.00	0.06
417	0	−60.52	Eq^{A}	8	0.30	−10.97	$L1_e$	0.00	0.44
202	0	−60.78	Eq^{A}	8	0.10	−9.94	Eq_e	0.00	0.31
518	0	−66.38	Eq	7	0.61	—	—	—	—
112	2	−66.39	Eq	7	0.00	−16.64	$L2_e$	0.25	0.25
215	0	−73.85	Eq	6	0.55	−3.74	$L1_e$	0.00	0.06
314	5	−78.06	Eq	5	0.52	−9.94	Eq_e	0.00	0.69
211	3	−79.14	Eq	5	0.00	−7.72	Eq_e	0.00	0.19
514	8	−85.98	Eq	4	0.00	—	—	—	—
406	2	−86.73	Eq	3	0.59	−6.03	$L1_j$	0.00	0.13
212	5	−96.62	$Eq^{\dagger+}$	1	0.00	−6.03	$L1_e$	0.00	0.88
506	0	−82.10	So	3	1.26**	—	—	—	—
304	5	−93.29	So^{+}	2	0.25	0.00	Eq_e	0.00	1.00
421	4	−96.78	$So^{\dagger}$	1	0.31	−10.59	Eq_e	0.00	0.38

Notes: A guesses-only type identifier superscripted † means the subject's estimated type was not significantly better than a random model of guesses ($\lambda = 0$, $\varepsilon \approx 1$) at the 5-percent (or 1-percent) level. A guesses-only type identifier superscripted + means the estimated type had lower likelihood than 12 or more pseudotypes, more than expected at random. A guesses-only type identifier superscripted A, B, C, D, or E indicates membership in a cluster. A guesses-only type identifier in bold indicates that the subject is classified as that type in Table 1, column 5, by the criteria stated in the text. An estimated λ superscripted ** (*) means that $\lambda = 0$ is rejected at the 1-percent (5-percent) level. A type-style identifier subscripted *el* indicates that both styles have equal likelihoods and ξ. A search-only type-style identifier subscripted ‡ indicates that there are alternatives with different types and/or ξ: $L1_j$ for subjects 101 and 404; $L2_e$ and $L3_e$ for 318 and 204; $L3_e$ for 416 and 201; $L2_j$ for 113; $L1_e$ and $L3_{el}$ for 309; $L1_e$ and $L3_e$ for 108; $L1_e$ for 316, 407, 403, and 315; $L1_{el}$, $L3_{el}$ and Eq_e for 412 and 312; $L1_j$, $D2_{el}$ and So_e for 313; and $D1_e$ for 303. No search estimates are reported for subject 109, who had zero search compliance in eight or more games for every type.

And most subjects’ types can be more precisely identified by decisions and search than by decisions or search alone (Table 7B):

Table 7B—Subject-by-Subject Guesses-and-Search Econometric Type Estimates

ID	Guesses and search							
	ln L_t	ln L_g	ln L_s	k	exact	λ	ξ_H	ξ_M
513	—	—	—	—	—	—	—	—
118	−17.03	−9.62	−7.41	$L1_e$	15	1.85	0.88	0.06
101	−20.21	−10.27	−9.94	$L1_e^{‡‡}$	15	0.55	0.69	0.31
104	−20.37	−16.63	−3.74	$L1_e$	14	2.20	0.00	0.94
413	−23.84	−17.81	−6.03	$L1_j$	14	0.88	0.13	0.88
207	−17.96	−17.96	0.00	$L1_e$	14	0.42	1.00	0.00
216	−38.69	−25.41	−13.29	$L1_e$	13	1.06	0.31	0.63
402	−39.93	−30.93	−9.00	$L1_e$	12	5.65	0.00	0.75
418	−52.16	−42.23	−9.94	$L1_e$	10	21.22	0.00	0.69
301	−49.58	−45.84	−3.74	$L1_e^{D}$	10	0.00	0.06	0.94
508	—	—	—	—	—	—	—	—
308	−60.65	−47.34	−13.30	$L1_{ej}$	10	0.00	0.19	0.69
102	−57.57	−47.63	−9.94	$L1_e$	10	0.00	0.00	0.69
415	−107.28	−90.90	−16.38	$D1_e^{+}$	2	0.76	0.31	0.50
504	—	—	—	—	—	—	—	—
208	−65.37	−61.62	−3.74	$L1_j$	8	0.00	0.06	0.94
318	−66.36	−62.61	−3.74	$L1_e$	7	3.18	0.00	0.94
512	—	—	—	—	—	—	—	—
502	—	—	—	—	—	—	—	—
516	—	—	—	—	—	—	—	—
409	−84.18	−73.59	−10.59	$L1_j^{E}$	4	9.90	0.00	0.38
106	−85.75	−75.82	−9.94	$L1_j$	5	1.19	0.00	0.31
305	−85.92	−79.89	−6.03	$L1_e$	5	0.37	0.88	0.13
411	−86.61	−80.58	−6.03	$L1_e$	4	1.45	0.13	0.88
509	—	—	—	—	—	—	—	—
203	−94.49	−83.90	−10.59	$L1_e$	4	0.00	0.00	0.63
505	—	—	—	—	—	—	—	—
317	−90.32	−86.58	−3.74	$L1_e$	3	0.92	0.94	0.06
416	−90.48	−86.74	−3.74	$L1_e$	1	4.48	0.00	0.94
217	−97.71	−87.12	−10.59	$L1_e$	3	0.68	0.00	0.38
219	−95.04	−87.32	−7.72	$L1_e^{+}$	3	0.89	0.00	0.81
501	—	—	—	—	—	—	—	—
410	−96.90	−89.18	−7.72	$L1_{ej}$	2	1.53	0.00	0.19
510	—	—	—	—	—	—	—	—
420	−94.26	−90.52	−3.74	Eq_j^{+}	3	0.19	0.00	0.06
408	−95.74	−89.71	−6.03	$L1_e^{+}$	2	1.09	0.00	0.88
201	−94.00	−90.26	−3.74	$L1_e^{+}$	2	1.21	0.00	0.94
105	−102.56	−93.56	−9.00	Eq_e^{+}	2	0.11	0.25	0.75
103	−96.63	−90.61	−6.03	$L1_e^{+}$	2	1.12	0.00	0.13
213	−100.34	−96.60	−3.74	$L2_e^{+}$	0	0.62	0.94	0.00
515	—	—	—	—	—	—	—	—
113	−108.49	−98.86	−9.63	$L3_{ej}^{+}$	4	0.00	0.81	0.06
109	—	—	—	—	—	—	—	—
309	−9.94	0.00	−9.94	$L2_{ej}$	16	0.00	0.69	0.00
405	−14.40	0.00	−14.40	$L2_e$	16	0.00	0.63	0.25
206	−17.49	−10.07	−7.41	$L2_e$	15	0.79	0.88	0.06
209	−35.45	−25.51	−9.94	$L2_j$	13	0.96	0.69	0.31
108	−25.88	−25.88	0.00	$L2_e$	13	0.45	1.00	0.00
214	−41.33	−35.30	−6.03	$L2_e$	11	2.73	0.88	0.13
307	−48.51	−38.88	−9.63	$L2_j$	11	1.04	0.81	0.13
218	−53.84	−40.54	−13.30	$L2_j$	11	0.60	0.69	0.19
422	−61.82	−55.79	−6.03	$L2_e$	9	0.22	0.88	0.13
316	−72.26	−58.43	−13.84	$L2_j$	8	0.73	0.06	0.38
407	−67.00	−60.98	−6.03	$L2_e^{C}$	8	0.44	0.88	0.13
306	−75.68	−71.94	−3.74	$L1_j$	6	0.71	0.00	0.06
412	−69.43	−69.43	0.00	$L2_e$	6	1.05	1.00	0.00
205	−75.80	−75.80	0.00	$L1_e$	4	3.27	0.00	1.00

Table 7B—Continued

ID	Guesses and search							
	ln L_t	ln L_g	ln L_s	k	exact	λ	ξ_H	ξ_M
220	−76.70	−72.96	−3.74	$L2_e$	6	0.32	0.94	0.06
403	−86.91	−80.88	−6.03	Eq_j^{+}	4	0.84	0.00	0.13
517	—	—	—	—	—	—	—	—
503	—	—	—	—	—	—	—	—
414	−102.56	−92.62	−9.94	Eq_e^{+}	2	0.36	0.00	0.31
110	−107.03	−98.03	−9.00	$L1_j^{+}$	0	0.56	0.00	0.75
210	−68.44	−51.13	−17.32	$L2_e^{B}$	9	0.92	0.38	0.25
302	−71.14	−65.12	−6.03	Eq_e^{B}	7	1.11	0.00	0.13
507	—	—	—	—	—	—	—	—
313	−90.93	−84.90	−6.03	$L1_e^{‡‡E}$	3	3.28	0.00	0.88
312	−84.74	−81.00	−3.74	$L2_e$	1	1.37	0.94	0.06
204	−88.47	−88.47	0.00	$L1_e^{+E}$	3	1.59	0.00	1.00
115	−107.99	−98.05	−9.94	Eq_e^{+}	0	0.39	0.00	0.31
401	−104.35	−98.32	−6.03	Eq_j^{+}	0	0.32	0.00	0.13
310	−56.84	−41.69	−15.15	Eq_{ej}	11	0.00	0.13	0.31
315	−50.80	−41.80	−9.00	Eq_e	11	0.00	0.00	0.75
404	−63.69	−54.69	−9.00	Eq_e	9	0.03	0.00	0.75
303	−63.68	−59.93	−3.74	Eq_e	8	0.41	0.00	0.06
417	−73.80	−60.52	−13.29	Eq_e^{A}	8	0.30	0.31	0.63
202	−70.72	−60.78	−9.94	Eq_e^{A}	8	0.10	0.00	0.31
518	—	—	—	—	—	—	—	—
112	−106.23	−89.60	−16.64	$L2_e^{+}$	3	0.00	0.25	0.25
215	−81.57	−73.85	−7.72	Eq_e	6	0.55	0.00	0.19
314	−87.99	−78.06	−9.94	Eq_e	5	0.52	0.00	0.69
211	−86.86	−79.14	−7.72	Eq_e	5	0.00	0.00	0.19
514	—	—	—	—	—	—	—	—
406	−99.17	−86.73	−12.44	Eq_j	3	0.59	0.06	0.25
212	−104.34	−96.62	−7.72	Eq_e^{+}	1	0.00	0.00	0.81
506	—	—	—	—	—	—	—	—
304	−97.31	−97.31	0.00	Eq_e^{+}	1	0.00	0.00	1.00
421	−109.34	−98.38	−10.97	$L1_e^{+}$	0	0.43	0.00	0.56

Notes: A guesses-and-search type identifier superscripted + means the estimated type had lower likelihood than 12 or more pseudotypes, more than expected at random. A guesses-and-search type identifier superscripted A, B, C, D, or E indicates membership in a cluster. A guesses-and-search type identifier in bold indicates that the subject is classified as that type in Table 1, column 6, by the criteria stated in the text. An estimated λ superscripted ** (*) means that λ = 0 is rejected at the 1-percent (5-percent) level. ln L_t, ln L_g, and ln L_s refer to total, guesses-only, and search-only likelihoods. ln L_t refers to total guesses-and-search likelihood. A type-style identifier subscripted ej indicates that both styles have equal likelihoods and ξ_e. A guesses-and-search type-style identifier subscripted ‡‡ indicates that there are alternatives with different ξ_e: $L1_j$ for subjects 101 and 313. No search estimates are reported for subject 109, who had zero search compliance in eight or more games for every type.

Econometric puzzles regarding CGC's analysis of search

Are there better ways to do the search analysis econometrically?

Our search analysis has so far focused on the order of look-ups. How can we incorporate duration data while retaining order information?

Can we say more about types' cognitive difficulty using duration data?

To what extent can Baseline subjects' guess "errors" be explained by a more detailed analysis of search?

Can we separate the effects of training from the strategic-uncertainty-eliminating effects of robot treatments?

Conditional on style, how does search differ between Baseline subjects with clear fingerprints (*Equilibrium*, *L1*, *L2*, or *L3*) and successful R/TS subjects of same type?

(Baseline subjects with high compliance for some type are like robot *un*trained subjects, which don't usually exist because you can't tell robot subjects how they will be paid without teaching them how the robot works, and so training them. Thus we can separate the effects of training and strategic uncertainty, by comparing Baseline and R/TS subjects:

Either *Equilibrium* is natural with mixed targets, but untrained subjects don't see it; or *Equilibrium* is unnatural, and/or subjects don't believe even trained others will make *Equilibrium* guesses with mixed targets.)

Possible answers via search to puzzle A. What are those Baseline "*Equilibrium*" subjects really doing?

(i) Can we tell how Baseline *Equilibrium* subjects find equilibrium in games without mixed targets: best-response dynamics, equilibrium checking, iterated dominance, or something else that doesn't "work" with mixed targets?

The absence of Baseline *Dk* subjects suggests that they are not using iterated dominance.

Best-response dynamics, perhaps truncated after 1-2 rounds, seems more likely.

Can check by refining characterization of *Equilibrium* search and redoing the searchmetrics, separately with and without mixed targets.

(At the very end of these slides is a refined characterization of *Equilibrium* search.)

(ii) Is there any difference in Baseline *Equilibrium* subjects' search patterns in games with and without mixed targets? If so, how does the difference compare to the differences for *L1*, *L2*, or *L3* subjects?

(Our 20 Baseline apparent *L1* subjects' compliance with *L1* guesses is almost the same with and without mixed targets (CGC's Figure 1, below), unsurprisingly because the distinction is irrelevant to *L1*.

But our 12 apparent *L2* and 3 apparent *L3* (CGC's Figures 2-3, below) subjects' compliance with their apparent types' guesses is lower with mixed targets. This is curious, because for *L2* and *L3*, unlike for *Equilibrium*, games with mixed targets require no deeper understanding.)

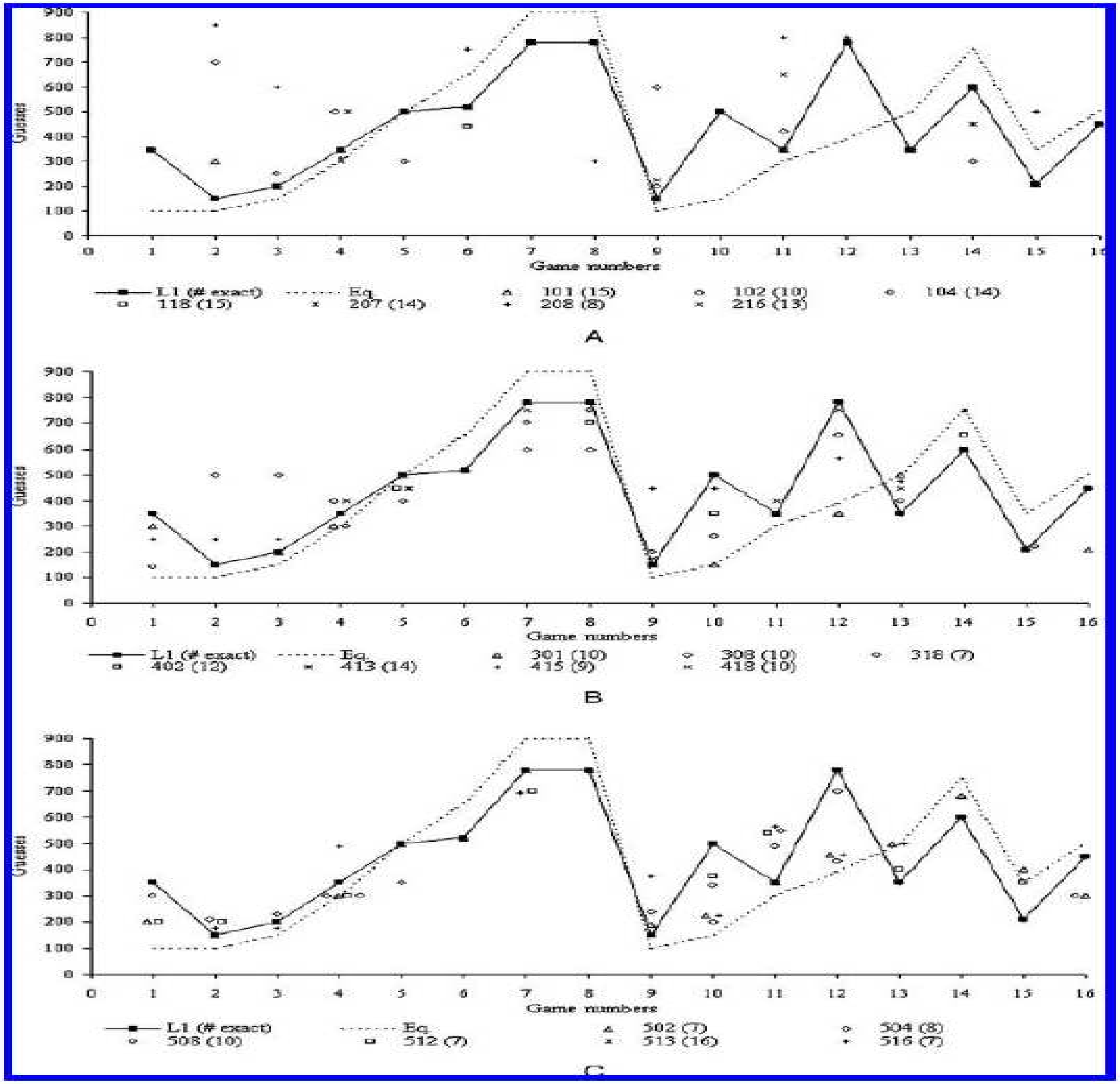


FIGURE 1. "FINGERPRINTS" OF 20 APPARENT *L1* SUBJECTS

Notes: Only deviations from *L1*'s guesses are shown. Of these subjects' 320 guesses, 216 (68 percent) were exact *L1* guesses.

CGC's Figure 1

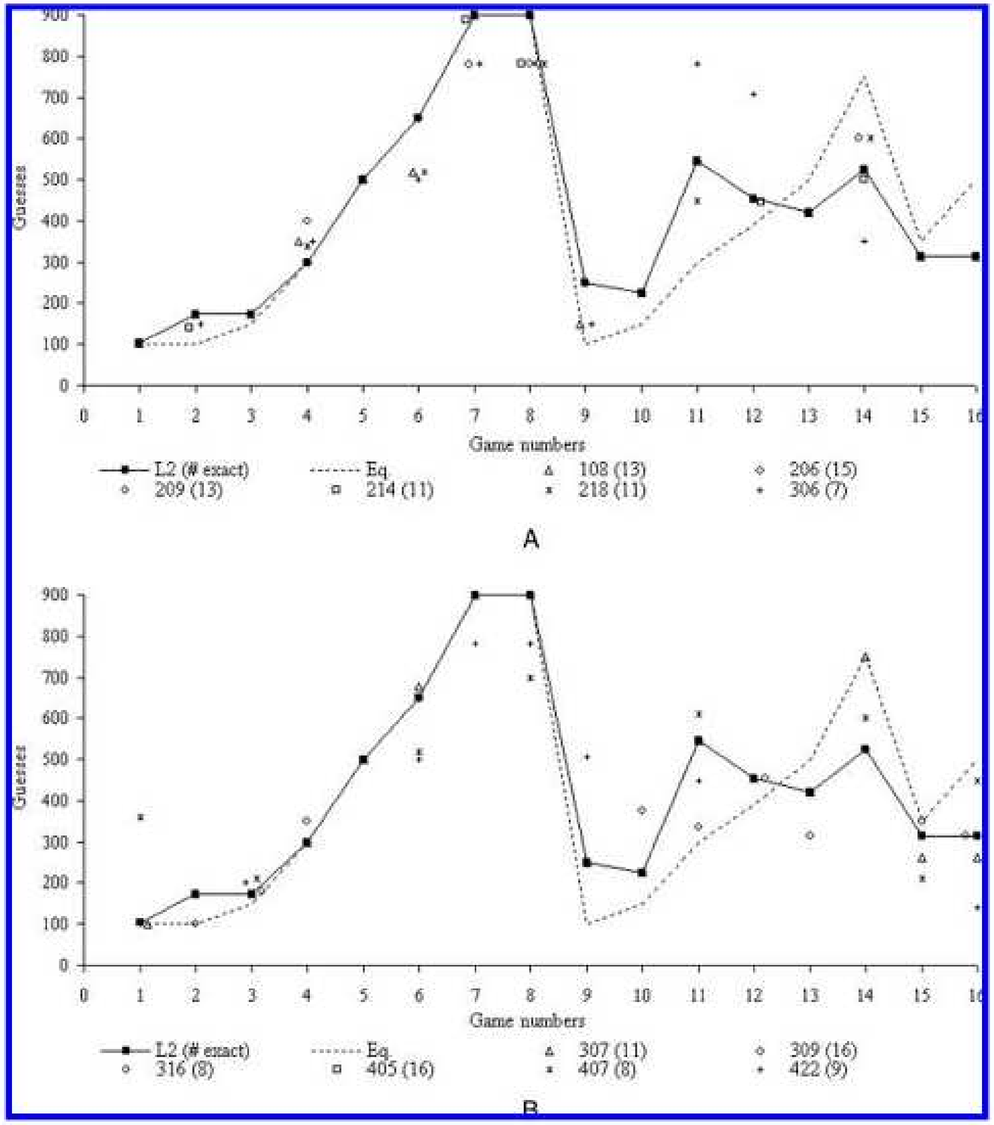


FIGURE 2. "FINGERPRINTS" OF 12 APPARENT *L2* SUBJECTS

Notes: Only deviations from *L2*'s guesses are shown. Of these subjects' 192 guesses, 138 (72 percent) were exact *L2* guesses.

CGC's Figure 2

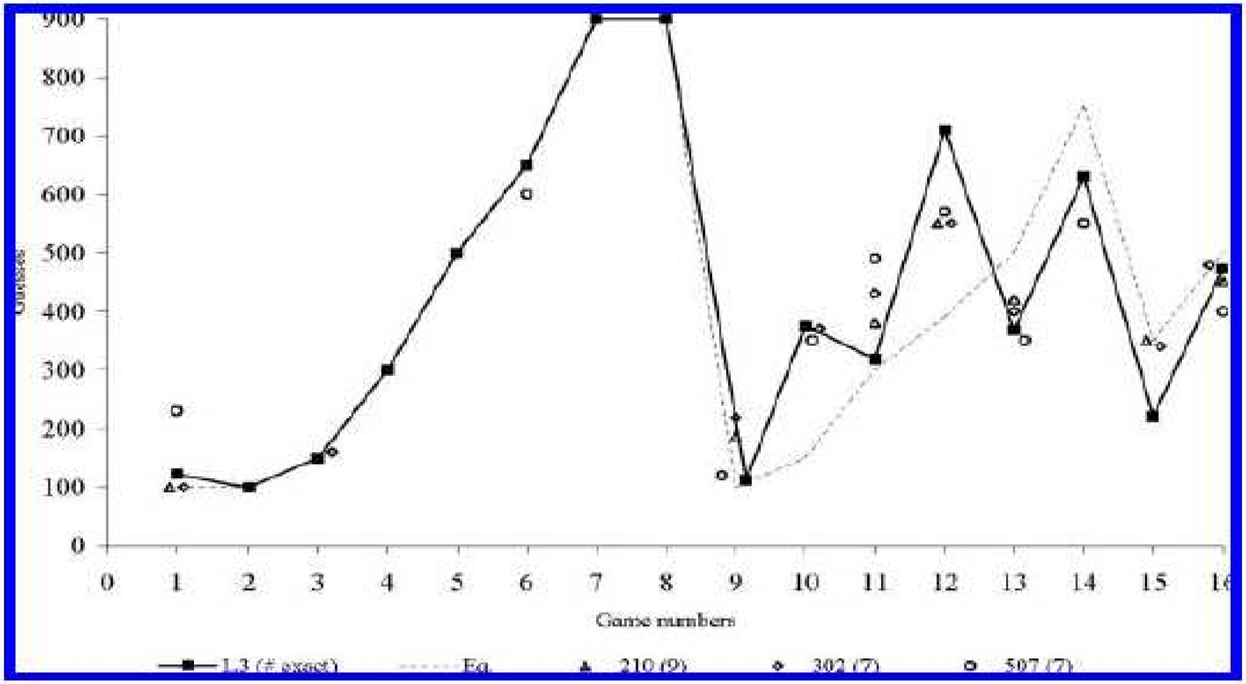


FIGURE 3. "FINGERPRINTS" OF THREE APPARENT *L3* SUBJECTS

Notes: Only deviations from *L3*'s guesses are shown. Of these subjects' 48 guesses, 23 (48 percent) were exact *L3* guesses.

CGC's Figure 3

(iii) Can we tell how R/TS *Equilibrium* subjects with high compliance manage to find their *Equilibrium* guesses even with mixed targets? How does their search in those games differ from Baseline *Equilibrium* subjects' search?

CGC strove to make the R/TS *Equilibrium* training as neutral as possible, but something must come first.

CGC taught them equilibrium checking first, then best-response dynamics, then iterated dominance (some were taught only one method).

To the extent that subjects used one of those methods, it explains why they have equal compliance with mixed targets.

If subjects used something else, and it deviates from equilibrium in games with mixed targets, it might provide a clue to what CGC's Baseline *Equilibrium* subjects did.

Does it help to know which Understanding Test questions an R/TS *Equilibrium* subject missed?

R/TS *Equilibrium* subjects' exact compliance is sensitive to the method that subjects were taught.

These average rates are for exact compliance, and so are quite high.

R/TS Subjects' Exact Compliance according to Equilibrium Method

	Eq.(N/A)	*Eq.(A)*	*EF*	*BR*	*ID*	*EqC*
Number of subjects	**29**	**50**	**11**	**13**	**13**	**13**
% Compliance|Passed UT2	**70.3**	**78.4**	**88.1**	**86.1**	**62.5**	**85.1**
% Failed UT2	**19.4**	**27.5**	**0.0**	**0.0**	**27.8**	**51.9**

Possible answers to puzzle B. Why are *Lk* the only types other than *Equilibrium* with nonnegligible frequencies?

(i) Most R/TS subjects could reliably identify their type's guesses, even *Equilibrium* or *D2*.

These average rates are for exact compliance, and so are quite high.

Individual subjects' compliance was usually bimodal within type, on very high and very low.

R/TS Subjects' Exact Compliance with Assigned Type's Guesses and Duration

	L1	***L2***	***L3***	***D1***	***D2***	***Eq.(N/A)***
Number of subjects	25	27	18	30	19	29
% Compliance|Passed UT2	80.0	91.0	84.7	62.1	56.6	70.3
% Failed UT2	0.0	0.0	0.0	3.2	5.0	19.4
Duration (seconds)	45.4	54.9	79.2	77	120.5	96.3

(ii) But there are noticeable signs of differences in difficulty across types:

(a) No one ever failed an *Lk* Understanding Test, while some failed the *Dk* and many failed the *Equilibrium* Understanding Tests.

(b) For those who passed, compliance was highest for *Lk* types, then *Equilibrium*, then *Dk*. This suggests that *Dk* is harder than *Equilibrium*, but could be an artifact of more stringent screening of the *Equilibrium* Test.

(c) Among *Lk* and *Dk* types, compliance was higher for lower *k* as expected, except *L1* was lower than *L2* or *L3* compliance.

(We suspect that this is because *L1* best responds to a random *L0* robot, which some subjects think they can outguess; *L2* and *L3* best respond to a deterministic *L1* or *L2* robot, which doesn't invite "gambling" behavior.)

(d) Remarkably, 7 of 19 R/TS *D1* subjects passed the *D1* Understanding Test, in which *L2* answers are wrong; and then "morphed" into *L2*s when making their guesses, significantly reducing their earnings (next slide).

(Recall that it is *L2* that is *D1*'s cousin.)

For example R/TS *D1* subject 804 made 16 exact *L2* (and so only 3 exact *D1*) guesses. Her/his search also suggests *L2* rather than *D1* thinking.

Table 10.2. Selected Robot/Trained Subjects' Information Searches.

Subject	Type/Alt[a]	Game 1[b]	Game 2[b]
804	D1 (3)/L2 (16)	1543465213	5151353654623

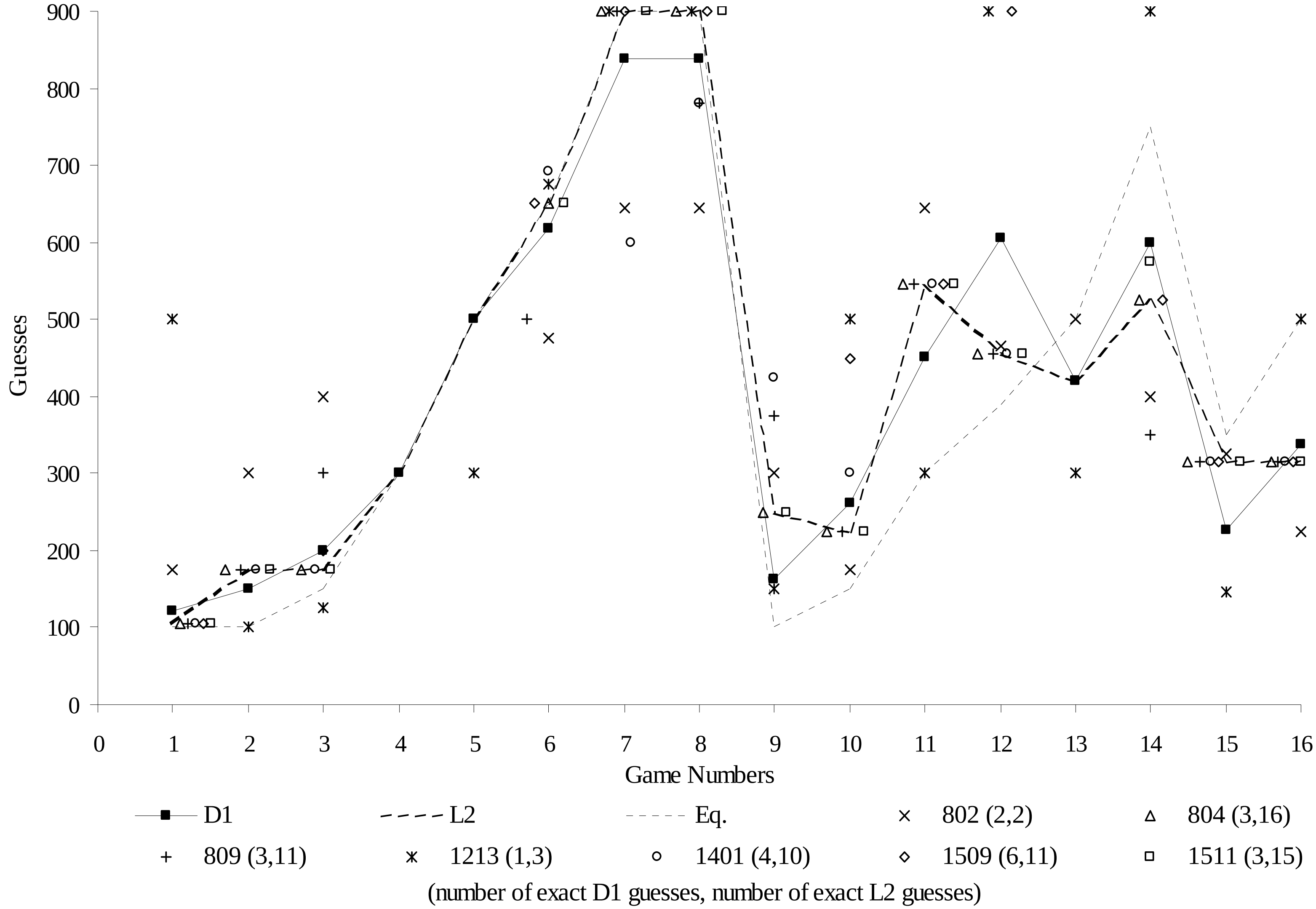
L2	{([1,3],5),4,6,2}
D1	{(4,[5,1], (6,[5,3]),2}

This kind of morphing, in this direction, is the only kind of morphing that occurred: compelling evidence that *Dk* types are unnatural.

However, a comparison of *Lk*'s and *Dk-1*'s search and storage requirements may add something, as *Dk-1* needs more memory than *Lk*.

Fingerprints of 7 R/TS Subjects who morphed from D1 to L2

(only deviations from D1's guesses are shown)



Aside: Refined characterization of *Equilibrium* search

Equilibrium's ideal guess can be identified by (1) evaluating a formula, (2) equilibrium-checking, (3) iterated dominance, or (4) best-response dynamics.

(1) Two ways to evaluate a formula: using *Equilibrium*’s ideal guess, or using Observation 1’s proxy for *Equilibrium*’s ideal guess.

Because they are logically related, our theory cannot distinguish them. The latter is less stringent, and yields requirements:

(1) $\{[p^i,p^j],a^j\} \equiv \{[2, 5], 4\}$ if $p^ip^j < 1$ or $\{[p^i,p^j],b^j\} \equiv \{[2, 5], 6\}$ if $p^ip^j > 1$.

(2) Equilibrium-checking’s requirements are almost the same, usually requiring both of the partner’s limits but excluding one in some cases, depending on luck.

I omit the requirements here, noting only that this method also requires $[p^i,p^j]$.

(3) Iterated dominance we assume requires one or more complete rounds, stopping when there is a clear up-or-down direction in which dominance eliminates guesses, enough to guess whether the equilibrium is High or Low.

Once the required rounds are completed, the player can use CGC's Observation 1's proxy for *Equilibrium*'s ideal guess; this adds a p^i times either a^j (Low equilibrium) or b^j (High) to his sequence.

As it happens, the search requirements for k rounds are independent of k; thus, the search requirements for iterated dominance are like CGC's characterization for *D2* (*D2*, not *D1*, because unlike *D1*, a k-round iterated-dominance player must delete k rounds of dominated guesses for himself too).

(3) $\{(a^i, [p^i, a^j]), (b^i, [p^i, b^j]), (a^j, [p^j, a^i]), (b^j, [p^j, b^i]), p^j, p^i\}$
$\equiv \{(1, [2, 4]), (3, [2, 6]), (4, [5, 1]), (6, [5, 3]), 5, 2\}.$

(4) For best-response dynamics we assume the subject does only one complete round: that is, starting with a trial guess for one player, best-responding for the other, and then best-responding back for the first player.

We also assume the subject can infer from whether the iterated best response goes up or down (if it changes) whether equilibrium is High or Low.

(4) {([a^i,p^j] or [b^i,p^j] or [a^j,p^i] or [(b^j,p^i]), p^i,p^j,(all but at most one of a^i,b^i,a^j, and b^j)}.

The main difference among *Equilibrium* methods is that methods 1 and 2 have a [p^i, p^j] requirement and methods 3 and 4 do not.

We know from the absence of Baseline *Dk* subjects in CGC’s guesses-and-search estimates that method 3’s requirements don’t fit the data well.

Its also seems, from the data, that $[p^i, p^j]$ are comparatively rare for Baseline apparent *Equilibrium* subjects, and even for R/TS *Equilibrium* subjects.

Thus searchmetrics may favor best-response dynamics, truncated 1-2 rounds.

(CGC strove to make the R/TS *Equilibrium* training as neutral as possible, but something must come first. A subset of the R/TS subjects were taught equilibrium-checking first, then best-response dynamics, then iterated dominance; another subset was taught only one of the methods. To the extent that they used one of those methods, it explains why they have equal compliance with and without mixed targets. If they used something else that deviates from equilibrium with mixed targets, it might be a clue to what Baseline *Equilibrium* subjects did.)

(End of aside.)